Alcohol Dehydrogenases from the thermophile Geobacillus thermoglucosidasius

Williams, Luke

Award date:
2016

Awarding institution:
University of Bath

Alcohol Dehydrogenases from the Thermophile *Geobacillus thermoglucosidasius*

Luke Peter Williams

A thesis submitted for the degree of
Doctor of Philosophy

University of Bath
Department of Chemistry
Centre for Doctoral Training in Sustainable Chemical Technologies

September 2016

CONTENTS

List of Figures iii
List of Tables vi
Acknowledgements vii
Abstract ix
Abbreviations xi
1. Introduction 1
1.1. Role of Biology in Chemical Synthesis 2
1.2. Whole Cell Versus Isolated Enzyme 9
1.3. Bioconversion-Chemistry-Engineering Interface 11
1.4. History and Taxonomy of *Geobacillus* Genus 16
1.5. Alcohol Dehydrogenases 18
1.6. Aims and Objectives of the Project 36
2. Investigations with Available Enzymes 37
2.1. ADH A 37
2.2. ADH B 53
2.3. ADH C 86
3. Investigations with Novel Enzymes 87
3.1. ADH D 87
3.2. ADH E 111
3.3. ADH F 128
3.4. ADH G 156
4. Discussion and Further Work 168
4.1. Results Summary 168
4.2. Native Role for Alcohol Dehydrogenases 169
4.3. Industrial Relevance of Enzymes 171
4.4. Additional Areas for Investigation 174
4.5. Summary of Implications from Sequence Aligments 175
5. Methods 176
5.1. Molecular Biology 176
5.2. Microbiology 179
5.3. Enzymology 179
5.4. Analytical Chemistry 183
6. References 184
7. Appendix One – Reference Data 202
7.1. Alternative Designations of Genes and Enzymes 202
7.2. List of All Primers 204
7.3. Full Sequences of Genes and Enzymes 205
7.4. Additional Sequences 210

LIST OF FIGURES

Figure 1 - Schemes of three generations of process research and development towards the manufacture of sitagliptin phosphate 9
Figure 2 - Monomer of RasADH 20
Figure 3 – Active site of RasADH in ribbon format 21
Figure 4 – Monomer A of TADH shown in ribbon format 23
Figure 5 - Model of phenylacetone in the active site of TADH 23
Figure 6 – Superimposition of active sites of TADH and TbADH 24
Figure 7 - Monomer of zmADH2 in ribbon form. 25
Figure 8 – Binding site of the iron metal in zmADH2 26
Figure 9 - Modelling of ethanol as substrate in the active site of zmADH2 26
Figure 10 - Structure of the ADH domain from the bifunctional *G. thermoglucosidasius* ADHE with front and back views 27
Figure 11 - Metal ion binding site of the ADH domain of the *G. thermoglucosidasius* ADHE 28
Figure 12 - Depiction of $NADP^+$ superimposed into the *G. thermoglucosidasius* ADH crystal structure of ADHE 28
Figure 13 - Ribbon diagrams of apo YtbE, apo YvgN and holo YvgN 30
Figure 14 - Structures of YvgN and YtbE showing the “safety belt” and catalytic tetrad 30
Figure 15 – Putative substrate binding site of YvgN 31
Figure 16 - SDS-PAGE gel of ADH A 37
Figure 17 – Graph of rate of reaction vs [S] using ADH A and ethanal 40
Figure 18 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and ethanal 40
Figure 19 - Graph of rate of reaction vs [S] using ADH A and propanal 41
Figure 20 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and propanal 41
Figure 21 - Graph of rate of reaction vs [S] using ADH A and butanal 42
Figure 22 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and butana 42
Figure 23 - Graph of rate of reaction vs [S] using ADH A and ethanal 44
Figure 24 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and ethanal 45
Figure 25 - Multiple Sequence Alignment (MSA) of *G. thermoglucosidasius* ADHE 49
Figure 26 - Multiple Sequence Alignment (MSA) of ADH A 50
Figure 27 - Butanol pathway under investigation 53
Figure 28 - SDS-PAGE gel of ADH B 54
Figure 29 - Graph of rate of reaction vs [S] using ADH B and 2-butanone. 57
Figure 30 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone 57
Figure 31 - Graph of rate of reaction vs [S] using ADH B and acetoin 58
Figure 32 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and acetoin 58
Figure 33 - Graph of rate of reaction vs [S] using ADH B and ethyl 4-chloroacetoacetate 59
Figure 34 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and ethyl 4-chloroacetoacetate 59
Figure 35 - Graph of rate of reaction vs [S] using ADH B and propanal 60
Figure 36 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and propanal 60
Figure 37 - Graph of rate of reaction vs [S] using ADH B and 2-propanone 61
Figure 38 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-propanone 61
Figure 39 - Graph of rate of reaction vs [S] using ADH B and 2-butanol 62
Figure 40 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanol 62
Figure 41 - SDS PAGE gel of Insoluble and Soluble Fractions subjected to sonication 64
Figure 42 - Graph of rate of reaction vs [S] using ADH B and ethanal 66
Figure 43 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and ethanal 66
Figure 44- Graph of rate of reaction vs [S] using ADH B and NAD^+ 67
Figure 45 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and NAD^+ 67
Figure 46 - Graph of rate of reaction vs [S] using ADH B and 2-butanone 68
Figure 47 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone 68
Figure 48 - Graph of rate of reaction vs [S] using ADH B and butanal 69
Figure 49 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and butanal. 69

Figure 50 - Graph of rate of reaction vs [S] using ADH B and 2-butanone. 70
Figure 51 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone 70
Figure 52 - Graph of rate of reaction vs [S] using ADH B and 2-pentanone 71
Figure 53 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-pentanone 71
Figure 54 - Graph of rate of reaction vs [S] using ADH B and pentanal 72
Figure 55 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and pentanal 72
Figure 56 - Graph of rate of reaction vs [S] using ADH B and 3-pentanone. 73
Figure 57 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 3-pentanone 73
Figure 58 - Graph of rate of reaction vs [S] using ADH B and 2,3-pentanedione 74
Figure 59 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2,3-pentanedione 74
Figure 60 - Graph of rate of reaction vs [S] using ADH B and 1-penten-3-one 75
Figure 61 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 1-penten-3-one 75
Figure 62 - Graph of rate of reaction vs [S] using ADH B and 3-methylbutanal 76
Figure 63 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 3-methylbutanal 76
Figure 64 – Graph of specific activities of ADH B recorded with ethanal in assays conducted after different periods of static incubation at 4 °C 80
Figure 65 - Multiple Sequence Alignment of ADH B 83
Figure 66 - SDS PAGE gel of ADH D 88
Figure 67 – Graph of rate of reaction vs [S] using ADH D and butanal. 92
Figure 68 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and butanal 92
Figure 69 - Graph of rate of reaction vs [S] using ADH D and 1-phenyl-1,2-propanedione 93
Figure 70 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 1-phenyl-1,2-propanedione 93
Figure 71 - Graph of rate of reaction vs [S] using ADH D and 4-chlorobenzaldehyde 94
Figure 72 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 4-chlorobenzaldehyde 94
Figure 73 - Graph of rate of reaction vs [S] using ADH D and NADPH 95
Figure 74 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and NADPH 95
Figure 75 - Graph of rate of reaction vs [S] using ADH D and 5-norbornene-2-carboxaldehyde 96
Figure 76 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 5-norbornene-2-carboxaldehyde in acetonitrile 96
Figure 77 - Graph of rate of reaction vs [S] using ADH D and ethyl 4-chloroacetoacetate 97
Figure 78 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and ethyl 4-chloroacetoacetate in acetonitrile 97
Figure 79 - Graph of rate of reaction vs [S] using ADH D and ethyl 2-oxo-4-phenylbutyrate. 98
Figure 80 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and ethyl 2-oxo-4-phenylbutyrate 98
Figure 81 - Thermostability data acquired at 70^{o}C. 100
Figure 82 - Optimal temperature study with ADH D 104
Figure 83 - Multiple Sequence Alignment of ADH D 108
Figure 84 - SDS PAGE gel of ADH E 111
Figure 85 - Graph of rate of reaction vs [S] using ADH E and propanal 114
Figure 86 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and propanal 114
Figure 87 - Graph of rate of reaction vs [S] using ADH E and methanal 115
Figure 88 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and methanal 115
Figure 89 - Graph of rate of reaction vs [S] using ADH E and ethanal 116
Figure 90 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and ethanal. 116
Figure 91 - Graph of rate of reaction vs [S] using ADH E and NADH 117
Figure 92- Hanes-Woolf plot ([S]/v vs [S]) for ADH E and NADH 117
Figure 93 - Graph of rate of reaction vs [S] using ADH E and NADPH. 118
Figure 94 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and NADPH 118
Figure 95 - Graph of rate of reaction vs [S] using ADH E and butanal 119
Figure 96 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and butanal 119
Figure 97 - Optimal Temperature study of ADH E 122
Figure 98 - Multiple Sequence Alignment of ADH E 125
Figure 99 - SDS PAGE gel of ADH F 128
Figure 100 - Graph of rate of reaction vs [S] using ADH F and 1-phenyl-1,2-propanedione 132
Figure 101 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 1-phenyl-1,2-propanedione 132

Figure 102 - Graph of rate of reaction vs [S] using ADH F and 1-phenyl-1,2-propanedione 133
Figure 103 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 1-phenyl-1,2-propanedione 133
Figure 104 - Graph of rate of reaction vs [S] using ADH F and furfural 134
Figure 105 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and furfural 134
Figure 106 - Graph of rate of reaction vs [S] using ADH F and furfural 135
Figure 107 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and furfural 135
Figure 108 - Graph of rate of reaction vs [S] using ADH F and 3,4-hexanedione 136
Figure 109 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 3,4-hexanedione 136
Figure 110 - Graph of rate of reaction vs [S] using ADH F and 3,4-hexanedione 137
Figure 111 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 3,4-hexanedione 137
Figure 112 - Graph of rate of reaction vs [S] using ADH F and 2,2,2-trifluoroacetophenone 138
Figure 113 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 2,2,2-trifluoroacetophenone 138
Figure 114 - Graph of rate of reaction vs [S] using ADH F and 2,2,2-trifluoroacetophenone 139
Figure 115 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 2,2,2-trifluoroacetophenone 139
Figure 116 - Graph of rate of reaction vs [S] using ADH F and ethyl 2-oxo-4-phenylbutyrate 140
Figure 117 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and ethyl 2-oxo-4-phenylbutyrate 140
Figure 118 - Graph of rate of reaction vs [S] using ADH F and ethyl 2-oxo-4-phenylbutyrate 141
Figure 119 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and ethyl 2-oxo-4-phenylbutyrate 141
Figure 120 - Graph of rate of reaction vs [S] using ADH F and NADPH 142
Figure 121 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and NADPH 142
Figure 122 - Graph of rate of reaction vs [S] using ADH F and NADH 143
Figure 123 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and NADH 143
Figure 124 - Optimal temperature study with ADH F 148
Figure 125 - Preliminary thermostability test, conducted at 60 ^{o}C 149
Figure 126 - Extended thermostability test at 60 ^{o}C using furfural 150
Figure 127 - Multiple Sequence Alignment of ADH F 153
Figure 128 - SDS PAGE gel of ADH G 156
Figure 129 - Primary screen of ADH G with NADPH as cofactor. 157
Figure 130 - Secondary screen with ADH G and NADPH as cofactor. 157
Figure 131 - Graph of rate of reaction vs [S] using ADH G and butanal. 158
Figure 132 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and butanal 158
Figure 133 - Graph of rate of reaction vs [S] using ADH G and octanal 159
Figure 134 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and octanal 159
Figure 135 - Graph of rate of reaction vs [S] using ADH G and 5-norbornene-2-carboxaldehyde 160
Figure 136 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and 5-norbornene-2-carboxaldehyde 160
Figure 137 - Graph of rate of reaction vs [S] using ADH G and NADPH 161
Figure 138 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and NADPH 161
Figure 139 - Thermostability test with ADH G, 60°C, NADPH. 163
Figure 140 - Optimal temperature study with ADH G 164
Figure 141 - Multiple Sequence Alignment of ADH G 166

LIST OF TABLES

Table 1 – Relative rate of reaction obtained using ADH A with 10µl substrate in a 1ml assay..... 38
Table 2 - Kinetic constants calculated For ADH A 43
Table 3 - Initial substrate screen with ADH B 55
Table 4 – Alcohol screen using ADH B 55
Table 5 - Summary of kinetics obtained with ADH B..... 63
Table 6 - Further Kinetics Data Summarised 77
Table 7 - Summary of supplementary kinetic analysis using the Hill Equation. 78
Table 8 – Summary of solvent resistance data obtained with ADH B and 2-butanone 79
Table 9 - Summary of secondary screen of substrates 81
Table 10 - Initial screen of ADH D..... 89
Table 11 - Primary substrate screen of ADH D 90
Table 12 - Secondary substrate screen of ADH D..... 90
Table 13 - Summary of kinetic data with ADH D. 99
Table 14 - Tertiary screen with ADH D (buffers) 101
Table 15 - Tertiary screen with ADH D (additives and solvents) 102
Table 16 - Tertiary screen with ADH D (ions) 103
Table 17 - Summary of GC/MS analysed assays..... 106
Table 18 - Primary screen for ADH E 112
Table 19 - Secondary screen with ADH E 113
Table 20 - Summary of kinetic data with ADH E..... 120
Table 21 - Solvent stability analysis of ADH E..... 121
Table 22 - Summary of GC/MS data obtained with ADH E..... 123
Table 23 - Primary screen of ADH F with both cofactors. 129
Table 24 - Secondary screen of ADH F with both cofactors 130
Table 25 - Summary of kinetic data with ADH F..... 144
Table 26 - Tertiary screen with ADH F (buffers) 145
Table 27 - Tertiary screen with ADH F (additives and solvents)..... 146
Table 28 - Tertiary screen with ADH F (ions)..... 148
Table 29 - Summary of GC/MS data with ADH F 151
Table 30 - Summary of kinetic data obtained with ADH G..... 162
Table 31 - Solvent screen, ADH G..... 162
Table 32 - Summary of end point assays analysed by GC/MS carried out with ADH G 164
Table 33 - Consolidated results of specific activities noted with all ADHs studied 169
Table 34 - Industrially relevant substrates that ADH D and ADH F have demonstrated activity with171
Table 35 – List of all successful primers used to isolate genes encoding putative ADHs..... 204

ACKNOWLEDGEMENTS

The journey to obtain a PhD is a long and convoluted path, full of blind corners, hidden obstacles and occasionally long desolate landscapes more akin to the surface of Mars than anything else. I would like to thank each and every person who has hammered in a sign at the side of the road, let me hitchhike another mile closer to my destination or kept me company along the way.

Firstly I would like to thank my cohort in the Centre for Doctoral Training in Sustainable Chemical Technologies. Not just for introducing me to the delights of red wine on our trip to Devon, although that has been a gift that keeps on giving throughout my PhD, but for all the camaraderie of my first year at Bath. Oh, and the coffee. *Definitely the coffee!* No problem, exam, dissertation or lab explosion could survive a nuclear strength coffee break brewed in the SusLab.

Whilst traveling down my road I have meandered, treating discipline boundaries with a callous disregard. One upside from this is meeting lots of people. Starting with chemistry, I would like to thank current and past members of the Bull group, both for the random discussions in the office and for all the assistance in the lab over the years.

Moving to biology and biochemistry, I would like to thank all of the Danson, Leak, Pudney and Mason groups for their wealth of assistance and esprit de corps. Whether putting the entire scientific world to rights in the early hours of the morning, exchanging quips in the mid-afternoon about how *our* project had the biggest problem at the moment, or just cleaning the lab in the evening, there was never a dull moment. No matter the time of day or night, there was almost always someone to bounce ideas off of, help with a problem, or shout at a particular piece of equipment or colony of *Geobacillus* that decided it wouldn't play ball that day.

In particular, I would like to thank Jon Extance, Chris Hills and Giannina Espina for their invaluable assistance in enzymology and for helping me get my project off the ground. Thanks also to Carolyn Williamson and Chris Vennard, who through sharp wit and technical skill have shown me the ways of microbiology, and even managed to get me to keep a lab bench clear. Occasionally. I will always remember that the lab seemed to have more spacecraft than spectrophotometers for some reason…

Special thanks go to my supervisors, to whom I am eternally grateful for all the help and support that they have provided me throughout my PhD. It may be a cliché; however, it remains true that I would not have got to where I am without them.

Throughout my PhD I have also toured a variety of strange worlds, whether they are set in the past, present or future. As with all journeys, this one had a beginning in a pub. Which exploded. From an exploding pub to shugenja by way of insurmountable fences, space stations, fireballs, and an inadvisably large supply of pizza. Thus I thank all those who have attempted to keep me sane by systematically destroying any semblance of sanity that I may have ever once possessed. Those who say that truth is stranger than fiction simply haven't experienced enough fiction.

Lastly I would like to thank my family for the assistance and encouragement that they have given me during my time at Bath and my entire life.

This thesis is dedicated to the memory of Florence “Angel” Baulsh.

A much loved and sorely missed matriarch.

There are many different paths running up to the mountaintop,

But everyone sees the same moon on the peak.

– Zen saying, trans. Soiku Shigematsu

ABSTRACT

This is an investigation into alcohol dehydrogenases (ADHs) from *Geobacillus thermoglucosidasius*. Eighteen ADHs have been studied, with seven taken for closer inspection. Characterisation was carried out to determine the industrial significance of these enzymes, starting with the substrate scope of the ADHs. The key results obtained are as follows:

ADH A is the alcohol dehydrogenase domain of the bifunctional ADHE enzyme (Extance, 2012; Extance et al., 2013). It has been determined that the substrate scope, whilst restricted to linear aliphatic aldehydes, extends at least to dodecanal. Also, with a specificity constant of 167 mM^{-1} min^{-1} it appears that ADH A could prefer butanal to shorter-chain aldehydes such as ethanal and propanal with specificity constants of 38 mM^{-1} min^{-1} and 35 mM^{-1} min^{-1}, respectively. Thus ADH A may have a preference for longer aldehydes than previously believed due to its native role in the production of ethanol from acetyl-coA.

ADH B was previously investigated for its potential role in the production of butanol. Here it was confirmed as an NADH-dependent ADH, with a substrate scope limited to five carbon length substrates and smaller, with residual activity with C6 substrates. ADH B demonstrated activity with ethyl 4-chloroacetoacetate, an intermediate in the production of statins. Further, an estimated half-life whilst stored at 4°C of 770 days; retention of 86% activity with 10vol% ethyl acetate and 92% activity with 10vol% acetonitrile; and a specific activity of 27 U mg^{-1} with 3M 2-butanone are all indications that ADH B is a potentially useful enzyme for industry.

The last enzyme to be previously investigated was ADH C, which in this work was confirmed to be an acetoin reductase with a very small substrate scope exclusively based around the acetoin motif, and therefore no further work was conducted.

ADH D and ADH F both have broad substrate scopes including the industrially-relevant substrates, 5-norbornene-2-carboxaldehyde, 1-phenyl-1,2-propanedione, ethyl 4-chloroacetoacetate and ethyl-2-oxo-4-phenylbutyrate. ADH D is an NADPH-dependent enzyme whereas ADH F can utilise both NADH and NADPH. Both enzymes are annotated as aldo-keto reductases, which is further indicated by multiple sequence alignment with the most similar available protein sequences and crystal structures. Thus, these two enzymes are the first aldo-keto reductases to be examined from moderate thermophiles, and are tentatively assigned in the AKR family as AKR6D1 and AKR5G4 respectively.

ADH D has a very low K_M (≤0.1 µM) with NADPH, giving a specificity constant of 2,800,000 mM^{-1} min^{-1}, substantially higher than any other noted. ADH D showed >80% activity from pH 5.0 - 8.0. The enzyme was resistant to solvents DMSO (at 5 vol%) and ethyl acetate, acetonitrile and cyclopentyl methyl ether (at 20vol%).

ADH F had the broadest substrate scope of any ADH tested, with 1-phenyl-1,2-propanedione the most preferred substrate with a K_M of 0.010 mM and a specificity constant of 54,000 mM^{-1} min^{-1}. It greatly preferred sodium phosphate at pH 7.0, as almost any deviation resulted in a substantial loss of activity. Activity of ≥70% was recorded in 5vol% DMSO, ethyl acetate, acetonitrile, cyclopentyl methyl ether and 50vol% hexane .

Both ADH D and F have optimal activities at 70 °C and both may have application in the biotechnology industry for the production of pharmaceutical intermediates and other high value chemicals.

ADH E acts solely as an aldehyde reductase, with V_{max} using NADH of 74, 331, 320 and 281 U mg^{-1} for methanal, ethanal, propanal and butanal, respectively. Activity with NADPH was limited (< 1% compared with NADH). Activity was also noted with higher aldehydes such as octanal and furfural.

ADH G is an NADPH-dependent ADH utilizing aldehydes only. It has an optimal temperature of 60°C with a half-life of under two hours at that temperature.

In conclusion, this thesis reports a feasibility study into the potential industrial use of specific enzymes for a variety of purposes ranging from the production of pharmaceutical intermediates to bioremediation. ADHs D and F are most likely to have use in the biotechnology industry, and ADHs B and E may be suitable for cofactor regeneration. ADH E may additionally be useful in the bioremediation industry. In addition, the anticipated biological significance of these enzymes is described.

ABBREVIATIONS

ADH – Alcohol Dehydrogenase

AKR – Aldo-Keto Reductase

BCR – Biocatalytic Cascade Reactions

DSP – Downstream Processing

IMAC – Immobilised Metal Affinity Chromatography

ISPR – *In Situ* Product Removal

MDR – Medium Chain Dehydrogenase/Reductase

MSA – Multiple Sequence Alignment

NAD^+ - Nicotinamide Adenine Dinucleotide

$NADP^+$ - Nicotinamide Adenine Dinucleotide Phosphate

NCBI – National Center for Biotechnology Information

PDB – Protein Databank

SDR – Short Chain Dehydrogenase/Reductase

STY – Space-Time Yield

SyPaB – Synthetic Pathway Transformations

Tris - Tris(hydroxymethyl)aminomethane [2-Amino-2-hydroxymethyl-propane-1,3-diol]

1. INTRODUCTION

With the current drive towards a sustainable chemical industry, it has become more frequent to consider biological solutions to problems previously considered the sole domain of chemistry (Clouthier and Pelletier, 2012). The derivation of chemicals from biorenewable sources, biocatalysis for the further processing, and bioremediation of waste streams are distinct elements that are all being pursued to a greater or lesser degree depending on their technological requirements. The fusion of these individual elements into existing processes is one approach; the *de novo* creation of entirely new processes is another (Vennestrøm et al., 2011).

By the application of new ways of thinking with traditional approaches, utilising old technologies with new materials and substrates, new routes to current chemical products can be discovered and innovative additional products can be found. The current cultural shift towards sustainability is a phase that is ending, giving way to the realisation that sustainability is not merely a sociological phenomenon, but by approaching technical problems in a holistic manner, it is possible to discover new solutions. There is a compelling business case, a compelling technical case and a compelling ecological case for a sustainable chemical industry. The only way a truly sustainable chemical industry can be created is by the use of biology, by replacing the carbon source in the industry with biologically derived materials rather than oil. This is not a simple or easy process; however, the concept of a biorefinery is important and a necessary one (Clark, 2007; Kamm and Kamm, 2007).

This will only become possible through the application of apparently disparate disciplines and fields; it is crucial to look forward and anticipate the required union of the biological and chemical sciences with their counterparts in the engineering disciplines. There is a plenitude of fields that could be useful (Murphy, 2011); however, one route is to focus on the enzymology and industrial biotechnology areas as critical mediators between the products needed and the source materials on offer (Hatti-Kaul et al., 2007; Wohlgemuth, 2009; Wenda et al., 2011; Philp et al., 2013).

Specifically alcohol dehydrogenases (ADHs) are enzymes that have been growing increasingly popular industrially, with examples being used across the chemical industry and in other fields (Musa and Phillips, 2011). This is because they are heavily studied and also because the products that they form are in high demand.

Additionally it is becoming evident that thermophilic organisms have something to offer the chemical industry. Previously considered useful only in techniques such as Polymerase Chain Reactions (PCR), thermophiles are gradually making progress as a major source of useful enzymes and as key species for further investigation (Niehaus et al., 1999).

Thus the literature will be reviewed with focus on the role of alcohol dehydrogenases from thermophilic organisms such as *Geobacillus thermoglucosidasius*, but with emphasis on the wider context within the fields of chemical synthesis and sustainability and the chemical industry. Examples will be drawn from a range of relevant industries, including the pharmaceutical, cosmetic, fine chemical, bioplastic, biofuel and water industries as representative of the extensive interconnections that exist between these areas.

1.1. ROLE OF BIOLOGY IN CHEMICAL SYNTHESIS

This section will consider the ways in which biology and associated disciplines have already become part of chemical synthesis, and the reasons behind this. A full analysis of sustainable development is outside the scope of this review and therefore analysis will be constrained to a comparative analysis of potential effects of biology, biotechnology and associated disciplines to the sustainability of the chemical industry.

1.1.1. SUSTAINABLE CHEMISTRY

Sustainable chemistry, or green chemistry, is the development of new chemical processes that are inherently safer, produce less waste, and are environmentally benign. Starting in the early 1990s, this field has become increasingly well supported through academic interest, industrial support and political will (Anastas and Kirchhoff, 2002; Wohlgemuth, 2009; Dunn, 2011). There is still debate on which metrics to use for measuring the "greenness" of a process (Jiménez-González et al., 2012); however, various methods such as atom economy and E factor are among those considered valid. In the case of atom economy, a process is considered better if a greater proportion of the atoms used as reactants end up in the final product rather than as waste. Importantly, this measure ignores solvents and additional reagents not directly involved with the reaction, whereas the E factor deliberately includes all waste streams and is defined as indicated in Equation 1 (Sheldon, 2012). Deliberation continues as to whether water should be included in this calculation, and typically is only included in the pharmaceutical industry.

$$E = \frac{Raw\ Materials - Desired\ Product}{Total\ Product}$$

Equation 1 - E factor equation, typically calculated on a mass basis, giving a value of kg waste per kg product. Typical values range from 1-5 kg waste per kg product for bulk chemicals through to 20-100kg waste kg^{-1} product or higher in the pharmaceutical industry (Sheldon, 2012).

There are various reasons behind the move towards green chemistry. Toxicity of waste streams is one particular driver, with the chemical industry producing voluminous quantities of waste that is difficult to dispose of. Even relatively benign materials such as sodium chloride can become major problems when considering industrial processes, as tons of such materials can still be problematic. Conversely, waste streams containing heavy metals can become an issue with very small quantities.

Another pressure on the chemical industry is climate change, and therefore the carbon footprint of the industry and the effects of emissions upon the atmosphere are coming under increasing scrutiny with public perception of climate change becoming more prevalent. Tied into this is the decreased reliance upon fossil fuels, either for energy or as raw materials, and this, *de facto,* encourages the industry to consider biological methods of obtaining raw materials.

A third area of consideration is cost and efficiency. As sustainability inherently considers cost and economic reasons to be integral to the concept, then it has become increasingly difficult for industry to marginalise sustainable chemistry as something alien and incomprehensible. Even if a company solely considers processes based upon financial considerations, green chemistry provides solutions which are intrinsically better in this regard.

Finally, as with all areas of industry, regulation is king. All the previously mentioned themes are emphasised with the development of regulations that either stimulate industry to change, or harshly encourage process development. In particular, the REACH regulations (Registration, Evaluation, Authorisation and restriction of CHemicals) instigated by the European Union and their US counterparts the Toxic Substances Control Act are particularly all-encompassing and compel the chemical industry to a higher standard than previously undertaken in terms of monitoring production and storage of toxic chemicals (Clark, 2007; Philp et al., 2013).

1.1.2. PRINCIPLES OF GREEN CHEMISTRY

The principles of green chemistry have been promulgated as follows (Anastas and Eghbali, 2010):

1. **Prevention.** It is better to prevent waste than to treat or clean up waste after it is formed.
2. **Atom Economy.** Synthetic methods should be designed to maximise the incorporation of all materials used in the process into the final product.
3. **Less Hazardous Chemical Synthesis.** Whenever practicible, synthetic methodologies should be designed to use and generate substances that pose little or no toxicity to human health and the environment.
4. **Designing Safer Chemicals.** Chemical products should be designed to preserve efficacy of the function while reducing toxicity.
5. **Safer Solvents and Auxiliaries.** The use of auxiliary substances (e.g. solvents, separation agents, etc.) should be made unnecessary whenever possible, and when used, innocuous.
6. **Design for Energy Efficiency.** Energy requirements of chemical processes should be recognised for their environmental and economic impacts and should be minimised. If possible, synthetic methods should be conducted at ambient temperature and pressure.
7. **Use of Renewable Feedstocks.** A raw material or feedstock should be renewable rather than depleting whenever technically and economically practicable.
8. **Reduce Derivatives.** Unnecessary derivatisation (use of blocking groups, protection/deprotection, temporary modification of physical/chemical processes) should be minimised or avoided if possible, because such steps require additional reagents and can generate waste.
9. **Catalysis.** Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
10. **Design for Degradation.** Chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment.
11. **Real-Time Analysis for Pollution Prevention.** Analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to formation of hazardous substances.
12. **Inherently Safer Chemistry for Accident Prevention.** Substances and the form of a substance used in a chemical process should be chosen to minimise the potential for chemical accidents, including releases, explosions and fires.

These principles can be arranged in groups loosely based around themes of safety, environmental protection and efficiency. In essence, the aim of any new chemical process should be to reduce waste, reduce toxicity and reduce safety concerns if it is designed with

reference to these principles. In keeping with other movements such as Quality by Design in the pharmaceutical industry (Blacker and Williams, 2011; Laird, 2011), these principles also encourage forward thinking and planning new processes carefully by considering alternatives wherever possible.

Another important aspect is put forward by Principle 10, the concept that not only should a chemical product be designed in conjunction with an efficient process, but that at the end of their usefulness they are designed to be easily degradable or perhaps otherwise used. The idea that a product beyond its useful lifespan can be utilised in a new capacity or constructed in such a manner as to allow easier recyclability or reusability is a relatively old idea only recently coming to the fore (Ellen MacArthur Foundation, 2015; Kiser, 2016; Stahel, 2016).

There are several fields of study that incorporate these ideas. Life-cycle assessments are a tool used to evaluate the sustainability of a given process, and these historically used only to incorporate the effects of production, called cradle-to-gate. More frequently they now consider cradle-to-grave, analysing the effects of transportation, sale, use and disposal of the product in question. They may even go cradle-to-cradle, which follows the waste materials of the product until the point they are utilised in new products (Hatti-Kaul et al., 2007). Industrial ecology is the study of the flow of materials and energy in industrial processes (Garner and Keoleian, 1995). Using ecosystems as a model, different feedstocks and routes can be compared for their efficiency and safety. These flows link different processes together, perhaps from disparate industries, in order to effectively absorb each component of a process so that there is no waste (Jacquet et al., 2015). This is termed the circular economy, whereby the products from a process are not simply discarded at the end of their useful lifetime, as in the linear economy. Central to this transformative approach is the concept of designing processes to be reversible and products to be easily degraded or reused, as implied by Principle 10 above (Stahel, 2016). One of the goals is to ensure that all flows from industrial processes are ultimately able to be discharged into environmental processes with no noticeable effect.

Obviously not every process is going to be able to maximise conformity to each of the principles of green chemistry or be integrated into a circular economy. There has also been debate as to whether a circular economy can ever realistically be obtained, with thermodynamic and social restraints present (Bilitewski, 2012; De Man and Friege, 2016). That said, when used as inspiration, it is possible that improvements will be seen on a sector or reaction basis.

1.1.3. SANDESTIN DECLARATION AND GREEN ENGINEERING

In July 2001 at the Sandestin Resort in Florida, the first conference on "Green Engineering – Defining the Principles" was held, leading to the attendees establishing what became known as the *Sandestin Declaration*. This was a set of principles of Green Engineering, replicating and extending the Principles of Green Chemistry to the engineering fields (Abraham and Nguyen, 2003). These draft principles were refined and finalised to read thus (Anastas and Zimmerman, 2003):

1. Designers need to strive to ensure that all material and energy inputs and outputs are as inherently non-hazardous as possible.
2. It is better to prevent waste than to treat or clean up waste after it is formed.
3. Separation and purification operations should be designed to minimise energy consumption and materials use.

4. Products, processes and systems should be designed to maximise mass, energy, space and time efficiency.
5. Products, processes and systems should be "output pulled" rather than "input pushed" through the use of energy and materials.
6. Embedded entropy and complexity must be viewed as an investment when making design choices on recycle, reuse or beneficial disposition.
7. Targeted durability, not immortality, should be a design goal.
8. Design for unnecessary capacity or capability (e.g. "one size fits all") solutions should be considered a design flaw.
9. Material diversity in multicomponent products should be minimised to promote disassembly and value retention.
10. Design of products, processes and systems must include integration and interconnectivity with available energy and materials flows.
11. Products, processes and systems should be designed for performance in a commercial "afterlife".
12. Material and energy inputs should be renewable rather than depleting.

Whilst the principles of both Green Chemistry and Green Engineering are broad, occasionally indistinct, and can be misinterpreted, their general emphasis is to consider not only the individual process under investigation, but also how that process will be used. Particularly in engineering, scale up is an important consideration and often overlooked at laboratory scale. Important too is a holistic viewpoint, with a life-cycle analysis of the system to prevent an improved process inadvertently causing bigger issues elsewhere in the system. This is particularly highlighted in Principle 6 of Green Engineering – embedded complexity refers both to the complexity of a given product but also the system in place to create the product. Embedded entropy in this case is similar, implying that a product with high complexity is likely to require substances with greater mass, more types of materials and higher entropy. This relates to the concept of a circular economy insofar as this principle means that these inherent characteristics of a product need to be taken into account when considering disposal or reuse. It also concerns the interactions of different disciplines during the lifetime of a product, which needs to incorporate a holistic approach involving both upstream and downstream perspectives. Scale-up is one aspect of this. An approach that may be very efficient at laboratory scale may suddenly be ineffective at pilot plant or larger scale. Principles 9 and 11 also refer to the end-of-life considerations of products, including the circular economy concept.

A simple example of this would be a chemical process with greater conversion, but would result in more expensive, energy intensive separation in order to effect a greater yield. This would be a situation where a "better" process may in fact cause a "worse" overall result and therefore needs to be considered carefully before implementation.

Though the majority of the principles often refer to overall analyses and can appear to be obvious, by way of example, consider principle 5 of Green Engineering. Typically in chemical reactions there is a surplus of at least one reactant, in order to drive the reaction to completion. Of benefit to analysis at bench scale, this is simply untenable at industrial scale, for cost, efficiency and sustainability reasons. Therefore, the implication from the principle is to create an *in situ* product recovery (ISPR) system whereby one of the products is removed to invoke Le Châtelier's Principle from the product side rather than the reactant side.

There are differences between Green Chemistry and Green Engineering, which reflects the different emphases between the fields. Whereas the principles of Green Chemistry are

written to ensure that a given process is as technically perfect as possible, with as few steps, with as much safety, and as ecologically sound as possible, the principles of Green Engineering are more focussed on the same issues at the system scale. In both cases though it is likely that any given process or system will require more investment in terms of financing and time in advance of implementation instead of piecemeal after implementation when problems arise. The adage “a stitch in time saves nine” is the general premise behind these principles.

1.1.4. GENERAL APPLICABILITY OF BIOLOGY TO PRINCIPLES

It is of benefit to utilise biotechnology in chemical synthesis and broadly in the chemical industry as by using a biological approach to complex chemical issues the principles of both Green Chemistry and Green Engineering can be adhered to. In particular, bioconversions and biotransformations are inherently safer than their chemical counterparts in terms of chemical safety by a reduced use of heavy metals, organic solvents and moderation of reaction conditions. Biological safety will depend upon the nature of the transformation undertaken, whether viable whole cells are used, the species involved and the specifics of each process.

The biological approach can be considered advantageous in many ways, but is not a panacea (Clouthier and Pelletier, 2012). In terms of renewability, as enzymes derive from alternative carbon sources to oil, this decouples the carbon input from fossil fuels. Waste streams are less likely to contain heavy metals and particularly toxic chemicals with severely damaging environmental and human health effects, yet they are much more likely to be extremely dilute aqueous streams which pose issues of volume and of processing.

Energy efficiency is outstandingly complex to determine as it is likely that a biological process will be carried out at a lower temperature than an equivalent chemical process, which would seem to indicate greater energy efficiency in the biological process. Conversely, biological processes may take longer, require substantial amounts of cooling to maintain 37 ^{o}C if a mesophilic organism is used, may necessitate energetically costly methods of separation to remove water from the product, and may have a greater number of by-products requiring further processing. Therefore, it may not be immediately possible to designate a biological system or process to be superior to an existing chemical process.

One major advantage of utilising a biological approach is that existing waste streams may be utilised as feedstocks for new processes. Furthermore, waste streams from any biological process may have use elsewhere as a feed for another process. Particular examples of the usefulness of this would be in the production of food, for instance corn, or in the biofuel industry where glycerol is an unwanted by-product (Zhou et al., 2008). This leads to the concepts of a biorefinery (Clark et al., 2006; Kamm and Kamm, 2007), industrial ecology and the potential for a circular economy. Substantial amounts of biomass are produced worldwide each year (Vennestrøm et al., 2011), with the majority of it being burnt or otherwise discarded. Lignocellulosic material is one of the most heavily considered materials as a biorenewable resource, and the full utilisation of it would represent a seismic shift in industrial carbon flow, with huge numbers of chemicals potentially being produced from lignocellulose, including vanillin, furans, sugars, ethanol and other alcohols as well as various organic acids. These may be used for a variety of purposes: biofuels, bioplastics and intermediates for an assortment of chemical processes (Hatti-Kaul et al., 2007; Serrano-Ruiz et al., 2011; Azadi et al., 2012).

As previously mentioned, the utilisation of renewable feedstocks does not obligate subsequent processing also to be biological as the two approaches are independent of each other. Thus work is ongoing to integrate renewable feedstocks into existing chemical processes whilst

equally research is conducted into developing biological routes to products, irrespective of the origin of the starting materials (Ran et al., 2008; May, 2009; Ma et al., 2010). This leads to a situation whereby renewable feedstocks may be used in a grossly wasteful chemical process or where a fossil fuelled process is catalysed biologically in an equally wasteful manner. By creating enhanced processes, of either chemical or biological nature, then gradually the chemical industry will become more sustainable, even if an apparent patchwork of unit operations are combined to create entirely new platforms for processing.

Two of the most well recognised sustainable processes are the production of artemisinin, an anti-malarial drug, and sitagliptin, a treatment for type 2 diabetes. In the first case, artemisinin is traditionally obtained botanically, from *Artemisia annua*, a plant used in Chinese medicine. Due to the difficulties in extracting the product and purifying it, it was expensive, especially so for an anti-malarial drug where cost may be a particular concern. Production volume was also a major limiting factor, as the herb itself requires an extensive amount of manpower to grow and harvest, which combined with relatively low yields, explains the prices of up to $1600/kg (Hale et al., 2007). The new process was to determine the metabolic pathway for the production of artemisinic acid, a direct precursor to the active molecule, and insert this pathway into yeast or *Escherichia coli*, which can be readily grown and harvested using existing technologies. By incorporating genes from a wide range of species, a full pathway of 11 genes was inserted into *E. coli* in order to determine the optimal production method. The methods used in this example required substantial effort; however, they have provided a framework that can be used for a great range of similar chemicals derived from isoprenoid pathways (Keasling, 2008; Anthony et al., 2009; Tsuruta et al., 2009).

The final production method utilised yeast as the chassis organism to increase production efficiency and has been developed into an industrial scale process producing artemisinic acid, which can then be chemically catalysed to produce artemisinin. This part-biological and part-chemical synthetic route is termed a semi-synthetic production process, and is successful at the multi-ton scale (Paddon et al., 2013; Paddon and Keasling, 2014). Whilst scientifically speaking a landmark case of synthetic biology, in terms of sustainability this process has been marred with flaws in economic design and social acceptance in recent years. Combined, these issues have led Sanofi, the commercial producer, to suspend production entirely. Economically speaking, the semi-synthetic production is not profitable at the current market price for artemisinin. Although the price for botanically acquired artemisinin reached $1600/kg on occasion, it fluctuated wildly due to the lead time for growing and harvesting the herb. More recently the price had stabilised, and has been as low as $250/kg, substantially below the breakeven point of $350-$400 established upon market entry. It has also not helped that charity funding, which was critical to the research phase, has inevitably been withdrawn once a commercial process was developed (Peplow, 2013, 2016).

Socially speaking, the botanical production of artemisinin involves an entire farming community and decentralised production and refinement. This is the main reason for the fluctuating prices, as farmers choose whether to plant *A. annua* or another crop. If the prices increase due to shortages, then many farmers plant the herb, causing mass overproduction shortly after. This is one of the reasons why a semi-synthetic route was considered, to stabilise production volumes and prices. There has been concern that the introduction of a commercial-scale, semi-synthetic process would have collapsed the botanical market, driving up prices for artemisinin and ruining the livelihood of the farmers who rely upon the crop. Instead, the above economic concerns have reduced the semi-synthetic method almost to a footnote (Peplow, 2013, 2016). Overall the sustainability credentials of this process are severely undermined due to both social and economic failures. Although the original goals of the project were admirable – to increase production of artemisinin to allow for mass

production of combined anti-malarial treatments and ultimately to save lives - semi-synthetic production of artemisinin may prove to be a cautionary tale on how not to commercialise synthetic biology triumphs.

The second example, sitagliptin manufacture, went through three or technically four iterations in total before the process was accepted. The full scheme is provided in Figure 1; however, in brief, the first generation process was an entirely chemocatalytic process plagued with substantial waste issues. An alternative chemocatalytic process was considered, the second generation process, which contained a three-step one-pot reaction to dehydrositagliptin (**12** in Figure 1), alternative starting materials to improve the atom economy, utilisation of effective catalyst recycling, and overall a much greater control than previous. Purification was also enhanced by employing crystallisation and filtration rather than chromatographic or distillation techniques. This process was scaled up successfully to production scale, having increased the overall yield from 52% to 65% and reduced the waste from 250kg to 50kg per kg of product.

Further work removed a reaction step, leading to an improved second-generation process, **(c)** in Figure 1, but this was still considered suboptimal due to the transition metal mediated step still in place. This was finally replaced with a heavily engineered transaminase to accommodate the substrate required, involving 27 mutations of the native enzyme. This final process led to a further increase in the yield (13%), productivity (53%) and reduction in the waste produced (19%). Typically, stereocontrol is a major reason for switching to an enzymatic route; however, this is not the overriding purpose in this particular case, as the final chemical process provided >99.9% *ee*. The issue was that, in achieving the necessary enantiomeric purity, a crystallisation step was needed and the use of transition metals required specialised equipment and demanded assiduous removal of the metal. By using the transaminase >99.95% *ee* was achieved, and this removed the necessity for a crystallisation step. Thus, whilst an increase in enantiomeric excess was achieved, it was a secondary effect driven by economics rather than technical necessity (Savile et al., 2010; Desai, 2011).

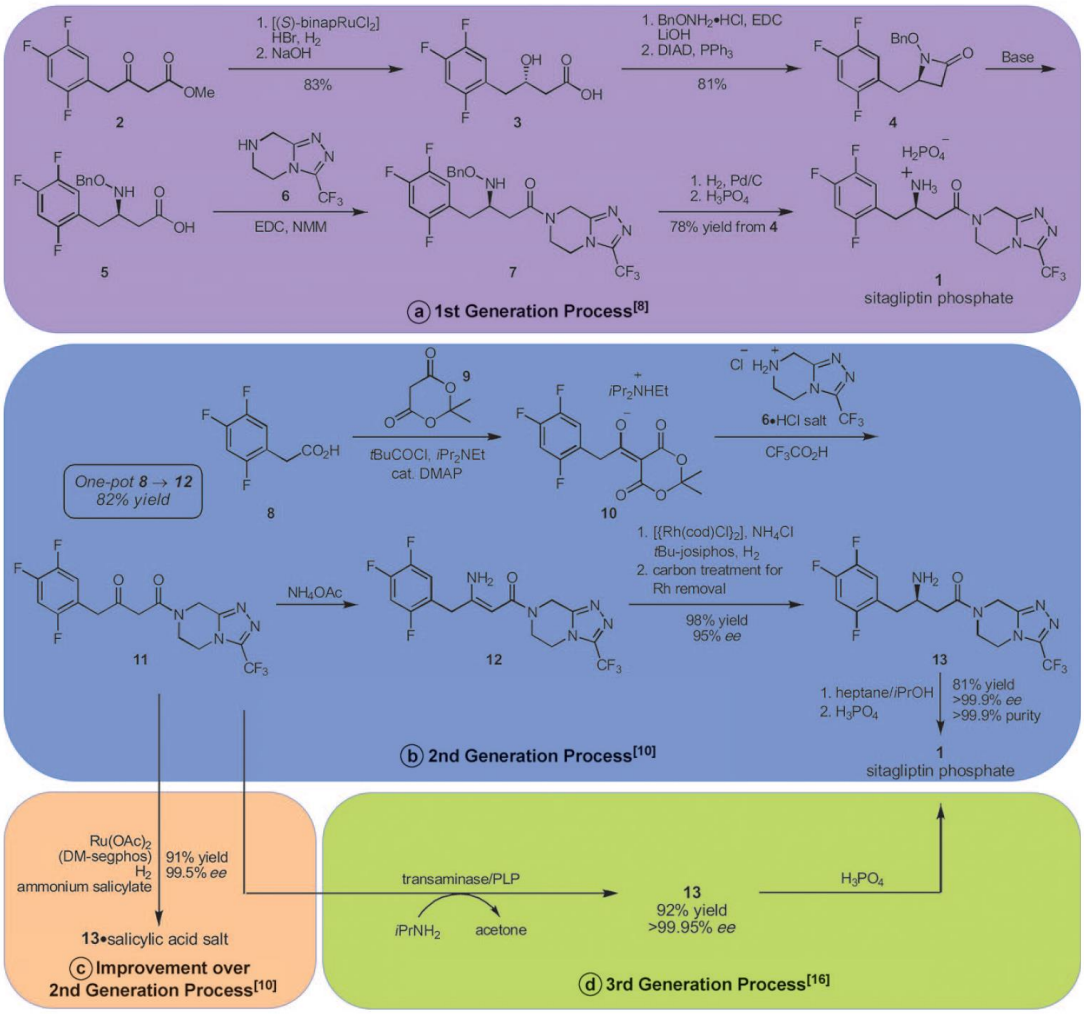


Figure 1 - Schemes of three generations of process research and development towards the manufacture of sitagliptin phosphate. Bn = benzyl, binap = 2,2'-bis(diphenylphosphino)-1,1'-binapthyl, cod = 1,5-cyclooctadiene, DIAD = diisopropyl azodicarboxylate, DMAP = 4-dimethylaminopyridine, NMM = N-methylmorpholine, segphos= (4,4'-bi-1,3-benzodioxole)-5,5'-diylbis(diphenylphosphine). Directly reproduced from Desai (2011).

1.2. WHOLE CELL VERSUS ISOLATED ENZYME

One of the common debates when considering an enzymatic reaction is whether to use whole cells or isolated enzymes to carry out the reaction, and this debate has evolved in recent years. Isolated enzymes are typically preferred to reduce side reactions, to ensure that the process undertaken is well understood, or to increase the space time yield (STY) of a process. They may also be used to improve downstream processing, which will be cleaner, or to operate in conditions that may be inimical to whole cells (Goldberg et al., 2007a). Something that is more rarely considered is the possibility of utilising multiple isolated enzymes for the purposes of creating a one pot multistep reaction. More traditionally a rule of thumb is used whereby if you need more than two or three enzymes for a specific reaction then it is considered best to attempt a whole cell approach. This usually refers to a single reaction that is sought, comprising multiple enzymatic steps, not an entire biosynthesis.

The new approach has been termed biocatalytic cascade reactions (BCR) (Guterl and Sieber, 2013) or synthetic pathway biotransformations (SyPaB) (Zhang et al., 2010), depending on the

emphasis of the work. In the case of biocatalytic cascade reactions the aim is to enable production of a required chemical product as a one pot reaction, using enzymes sequentially to biologically synthesise a product from cheaper and readily available substrates. In the case of synthetic pathway biotransformations, the aim is to replicate extant pathways in a cell-free environment or engineer superior pathways based upon natural equivalents, typically using a larger number of enzymes in both cases. Another viewpoint is that SyPaB concentrates on biosynthesis, whereas BCR is effectively biotransformation. Thus, whilst there may be considerable overlap between the two methods, the motivation is substantially divergent.

What is particularly interesting about these methods is that they are not entirely new. In chemistry the principle of domino reactions has been known for a long time (Tietze, 1996; Tietze and Modi, 2000), which are defined as "*a process involving two or more bond-forming transformations (usually C-C bonds) which take place under the same reaction conditions without adding additional reagents and catalysts, and in which the subsequent reactions result as a consequence of the functionality formed in the previous step*." Indeed, enzymatic domino reactions such as the formation of vitamin B_{12} are referenced (Tietze, 1996).

In addition to domino reactions, another key concept in organic synthetic chemistry is click chemistry. Inspired by nature, reactions deemed to be good enough to be designated click reactions are those which are essentially irreversible due to thermodynamic constraints, produce exceptional yields that require minimal purification or isolation, and frequently use water as the operating solvent. Whilst the reaction type or catalyst itself are not relevant, click chemistry is generally based around carbon chemistry, specifically the formation of larger molecules of interest from readily available small molecules in a modular method. Taking the biological production of proteins and DNA as stimulus, it is not hard to see how this has become an accepted synthetic philosophy that has grown alongside the popularity of green chemistry. Click chemistry was formally described by Kolb et al., (2001), with an update provided by Moses & Moorhouse, (2007).

From the examples above, it can be noted that one-pot reactions are considered the ideal solution in both chemistry and biology, for much the same reasons. Convenience and a predictable outcome are very important, but more than that a semi-batch process could be used. A semi-batch, or fed-batch, process is one in which substrates can be added sequentially over time and the entire product emptied at the end. Neither a batch nor continuous process, it is often used to maximise productivity from a single vessel (Azadi et al., 2012; Wettstein et al., 2012). It is also of interest on safety grounds. As the gradual addition of substrates adds a degree of control to the reaction, in the case of strongly exothermic reactions with high rates of reaction, gradual addition of a reactant can reduce the risk of a runaway reaction (Dale, 2011). This could also be helpful to maintain, for example, a stable pH in a biological process.

By definition a semi-batch process is not a domino reaction; however, neither SyPaB nor BCR are predicated on a purely batch environment. That said, BCR would seem to lend itself better towards a full batch process, with a view to replicating an entire biosynthetic pathway that is fully balanced in terms of cofactors and associated reagents (Guterl and Sieber, 2013).

In contrast to these cell free methods, the whole cell approach remains popular (Goldberg et al., 2007b) despite difficulties in maintaining cell viability and balancing this with a high productivity (Carvalho and Fonseca, 2002; Marques et al., 2010).

1.3. BIOCONVERSION-CHEMISTRY-ENGINEERING INTERFACE

It is important to acknowledge the different disciplines that could aid the progression towards a sustainable chemical industry and in particular the tools and techniques that could be deployed for an effective and sustainable synthesis platform. One term for this area of research is the Bioconversion-Chemistry-Engineering (BiCE) Interface as it is the intersection of all these areas which is critical to the area moving forward (Smith et al., 2010; University College London, 2014).

1.3.1. METABOLIC ENGINEERING, SYNTHETIC BIOLOGY AND CHEMICAL GENOMICS

Metabolic engineering and synthetic biology are both relatively young disciplines and already having a profound impact upon industrial biotechnology and the chemical industry. The former discipline is concerned with the determination of extant pathways in an organism, how they are organised and regulated, and how they may be improved by the addition, deletion or manipulation of specific genes in a pathway. The latter is focused on the *de novo* creation of pathways, frequently through the development of specific elements, in order to construct desired pathways and even organisms.

The two fields are frequently confused, and whilst there is overlap, they are distinct domains. An excellent differentiation between the two fields and their history is given by Stephanopoulos (2012). Alternative designations are provided by Jarboe et al. (2010) for metabolic engineering: "*the directed improvement of production, formation or cellular properties through the modification of specific biochemical reactions or the introduction of new ones with the use of recombinant technology*", and synthetic biology, "*the design and construction of new biological components, such as enzymes, genetic circuits, and cells, or the redesign of existing biological systems.*"

Metabolic engineering is already providing benefits in the biofuel arena (Lü et al., 2011), with the field directly leading to fourth generation biofuels, whereby algae are engineered to produce biofuels more efficiently. By determining exactly how algae produce fuels of interest, their genomes have been sequenced and metabolism determined in order to improve production of the biofuel of interest.

Further work in this area is the development of pharmaceuticals (Lu et al., 2009), other advanced biofuels (Atsumi and Liao, 2008; Zhang et al., 2011), bulk chemicals (Shin et al., 2013), enzymology and industrial biotechnology (Pirie et al., 2013),

Similarly, synthetic biology has been used in the production of biofuels (Connor and Atsumi, 2010; Jarboe et al., 2010; Kung et al., 2012), synthetic chemistry (Keasling, 2008; Carothers et al., 2009; Felnagle et al., 2012), production of natural products for drug discovery (Neumann and Neumann-Staubitz, 2010; Mitchell, 2011), regenerative medicine and other therapies (Burbelo et al., 2010; Ruder et al., 2011), production of reduced genome hosts (Gao et al., 2010) and elsewhere in the chemical industry (Du et al., 2011). Approaches include pathway engineering (Prather and Martin, 2008), integrated gene circuits (Nandagopal and Elowitz, 2011), DNA synthesis and organism level manipulations (Liang et al., 2011), creation of independent gene clusters that can be inserted as one unit (Fischbach and Voigt, 2010; Medema et al., 2010) and working with mixed populations of bacteria for a common purpose (Brenner et al., 2008).

Common to most discussions of synthetic biology is the concept of being able to approach a problem at multiple levels. Firstly at the DNA level, editing specific genes, gene regulation or

even creating entire clusters of genes regulated as necessary to provide optimal production of a given product or to restrict cofactor usage. Another option is at the protein level, by manipulation of enzyme active sites through site-directed mutagenesis, whether to improve affinity for a given substrate or cofactor, to improve a rate of reaction, or conversely to weaken performance of an undesired enzyme to remove surpluses of product. Alternatively, at organism level or above, it is possible to influence the entire genome by directed or adaptive evolution, or obligating an organism to respond to external stimulus, whether an additive or substrate in media or quorum sensing in one species or as part of a co-culture.

Taken to a logical extreme, synthetic biology is about the reduction of desired activities to as few elements as possible, and using these elements as required to recreate them in a more favourable environment. This is in contrast to metabolic engineering which is the epitome of holism, seeking to understand the entirety of a species and how it operates and how to manipulate it for the best advantage. Pushing this concept of metabolic engineering further, it is possible to conceive of testing an organism's response to given stimuli, not to merely encourage adaptive evolution, but to determine which genes are affected by the stimuli. This is precisely what has been done to identify new methods of action for antibiotic resistance (Roemer et al., 2011), but can be applied much more widely.

In the case of antibiotic resistance, a wide range of yeast strains were co-cultured with a potential antifungal agent and the relative proportions of different strains that grew were considered. Strains that grew better when overexpressing a particular gene, and worse when that gene was deleted, allow the method of action to be determined of that particular agent. In essence, the genes required for resistance are discovered. Therefore, if one gene were deleted from each strain, and in all strains tested every single non-essential gene was deleted, it could be determined from the relative proportions of survivors which genes had an influence upon the survival of the organism in the presence of a given inhibitory molecule.

This approach can be utilised to determine which genes encode for any other particular task, by altering the stimulus to instigate the response required. Essential genes and genes reacting to specific inhibitions can be discerned, and thus in principle any bioactive chemical can be used as a stimulus to uncover genes of interest. This could include chemicals such as alcohols, aldehydes and ketones, which can be toxic to cells at relatively low concentrations. This method is particularly important because it is typical for a given bioactive chemical to act on multiple gene products, making it difficult to determine all the potential interactions. This is the power of this particular method; it is possible to identify all the responses from a range of compounds without foreknowledge of each gene's activity or even if the chemicals being tested are even bioactive. This approach is termed chemical genomics and is another important field that needs to be considered.

Other areas of significance include the "omics" technologies, specifically genomics, proteomics and metabolomics, which cut across the areas previously mentioned (Dellomonaco et al., 2010; Yang et al., 2011).

1.3.2. CONVERGENCE AND THE BIOREFINERY

Convergence is a new research model based around the principle of multiple disciplines effectively merging together with the aim of defining research teams by the problem that they investigate rather than the discipline they work in. This takes traditional interdisciplinarity and multidisciplinarity a stage further, and indeed can incorporate disciplines and fields already considered interdisciplinary, for example synthetic biology.

Of particular note is the expectation that technologies, techniques and methods from quite disparate areas, from across the sciences and engineering fields, can be brought together in order to solve specific problems. It is proposed that this is already occurring in the area of healthcare, specifically biomedicine, where advances in fields such as optics, information technology and nanotechnology are affecting a traditionally life sciences led discipline. It is distinctively the application of ideas and techniques from one traditional discipline being applied to other disciplines, taking new approaches borrowed from entirely different areas of research and applying them as needed, modified to fit their new discipline. Given the first two revolutions in this area, molecular biology and genomics, this is termed the third revolution of the biomedical field (Sharp et al., 2011).

With the huge number of different disciplines and fields previously mentioned which may have relevance with regards to a sustainable chemical industry, it is plausible that convergence is necessary in this area, or perhaps in the more general area of sustainability, in order to accomplish the goal of converting the chemical industry into a clean and sustainable business (Philp et al., 2013).

To a certain extent the framing of disciplines and research methods can be a result of funding streams, thereby becoming as much of a policy decision as a purely technical choice. Thus it would be imperative that funding regimes, government and non-governmental groups could be persuaded of the necessity of such a change. Fortunately, this area is gaining traction, with a report by the Organisation for Economic Co-Operation and Development (OECD) linking industrial biotechnology and climate change as part of convergence. They recognise the importance of convergence as a vehicle for economic and scientific development: "*Due to the current clean technology boom and the reaction of the market and consumers, the development of "green technologies" that improve energy efficiency, are less polluting and less consuming, is being stimulated. An example is the so-called bio-inspired catalysis, at the boundary between biology and green chemistry."* (OECD Directorate of Science Technology and Industry, 2011)

Whilst this is an energising paradigm shift, it is clear that it is by no means brand new or a fast process, with convergence between biology and chemistry noted at least a decade ago (Kaul et al., 2004). That said, more disciplines than merely biological and informational sciences are now involved in the sustainable transformation of the chemical industry, with engineering, social sciences and humanities disciplines all being recruited, within sustainable development and climate mitigation as a whole (Kuiken, 2013).

To illustrate a potential route for convergence, one needs to look no further than the biorefinery. In traditional petrochemical purification it is necessary to fractionate crude oil to provide a wide range of different products such as petrol or bitumen. In this manner, as much of the crude oil is utilised as possible, providing as many products as possible. A similar approach can be taken with biomass as the feedstock, and using a small number of key intermediates as entry points for further processing into the full range of chemicals demanded by the chemical industry (Clark et al., 2006; Clark, 2007).

The Department of Energy (DoE) in the United States has generated a "Top Ten" biobased products that it believes are the most needed, and is stimulating science and engineering to find an acceptable sustainable method of producing these chemicals (Werpy et al., 2004). It is a point of interest that, common to petrochemicals, biofuels are a central part of the biorefinery concept (Bozell and Petersen, 2010). It is also crucial that whilst the raw materials noted are biorenewable in origin, the subsequent processing is not limited to biocatalysis, and well established chemocatalysis is also referenced heavily. The aim is to have key chemicals

that can provide the core of a new sustainable industry, not to convert everything to biocatalysis at all costs (Geboers et al., 2011; Serrano-Ruiz et al., 2011).

Whilst the DoE is the most comprehensive and well known of those who have specified requirements for a biorefinery, they are by no means alone (Patel et al., 2006; Kamm and Kamm, 2007). Whilst the DoE concentrated on a technologically independent list of candidate products, others have concentrated on specific technologies and the potential products that may be acquired from using specific feedstocks, e.g. lignocellulosic biomass (Dautzenberg et al., 2011; Wettstein et al., 2012) and glycerol (Johnson and Taconi, 2007), or even by species proposed to be central to biorefineries – *Bacillus subtilis* (Zhang and Zhang, 2010).

Even with the use of biorefineries there is still likely to be some waste. One of the final benefits of using biorenewable feedstocks and enzymatic transformations is that any by-products are highly likely to be useful in some manner to another organism. The challenge is to locate an economically viable solution for each by-product and this may require purification or separation of some sort. Whilst it is obviously preferable to prevent waste formation, as per the Principles of Green Chemistry, this is not always possible. Therefore the next best thing is to utilise all "waste" streams, recycling them into new systems. This will lead towards the development of a circular economy, whereby all waste streams from one unit operation become feedstocks for another unit operation. From a business perspective, this may require branching out into new fields unfamiliar and unrelated to the original project, which may necessitate careful thinking and consulting externally (Anastas and Eghbali, 2010).

The ultimate deployment of biorefinery technology and moving towards a holistic view of the chemical industry as envisaged by a circular economy will demand input from across the sciences, engineering, and beyond. Thus convergence is all but inevitable if this area of research is to proceed to an effective conclusion.

1.3.3. ADVANTAGES OF THERMOPHILES

Thermophiles, typically defined as organisms with optimal growing temperatures of 50 ^{o}C-80 ^{o}C, and hyperthermophiles, defined as organisms with an optimal growing temperature in excess of 80 ^{o}C, have a number of potential advantages over traditionally utilised mesophilic organisms. The most obvious advantage is that running a process at a higher temperature is liable to increase the rate of reaction, although this is mitigated somewhat by thermophilic enzymes being approximately as efficient as their mesophilic counterparts at their respective optimal temperatures. Higher temperatures are believed to improve operating conditions in situations such as the degradation of lignocellulosic biomass because of the frequent reliance on high temperature pretreatments of the biomass and because high temperatures also assist with the breakdown of the biomass (Frock and Kelly, 2012; Lin et al., 2014).

Broadly speaking, there are a very large number of enzymes already being considered from thermophilic organisms for industrial or research use (Willies et al., 2010; Chang and Yao, 2011; Bergquist et al., 2014; Elleuche et al., 2014, 2015; Littlechild, 2015; Siddiqui, 2015). This includes all three domains of life – *Bacteria*, *Archaea*, and even *Eukarya,* and ranges from the often used *Thermus aquaticus* DNA polymerase for molecular biology (Niehaus et al., 1999; Frock and Kelly, 2012) to polysaccharide degrading enzymes (Elleuche et al., 2015).

Archaea is the most thermophilic of the domains, with *Pyrococcus furiosus* among the most thermophilic species with an optimal growing temperature of 100 °C (Zeldes et al., 2015); the highest temperature at which life is known to survive is nearing 125 °C, which *Methanopyrus*

kandleri can withstand at a pressure of 200 bar (Elleuche et al., 2014). Individual enzymes may exceed even this, with enzyme activity being noted up to 130 °C (Siddiqui, 2015; Sammond et al., 2016). *Bacteria* is next in the thermophilicity table, with some organisms having optimal growth temperatures of up to 80 °C, including species such as *Thermotoga maritima* and *Thermus thermophilus* (Zeldes et al., 2015). *Eukarya* is the least thermophilic of the three, yet the most resilient of the fungi can reach up to 61 °C (Oliveira et al., 2015).

Many reactions are more advantageously conducted at high temperature, reducing viscosity, making reactions that were thermodynamically unfavourable possible and improving the range of options available. Examples include the production of fructose from corn syrup using glucose isomerase, hydrogen production and methane production (Zeldes et al., 2015). Additionally, solubility of many substrates increase with temperature (Alves et al., 2007; Sammond et al., 2016), which is another key benefit. This is not a constant benefit though, as there are examples where increasing temperature does not always increase the solubility (Stephenson et al., 1984; Stephenson and Stuart, 1986; Stephenson, 1993; Mączyski et al., 2008). Further, high temperature may degrade substrates and products, which must be considered.

One of the principal advantages thermophiles have is that thermophilic enzymes are more frequently resistant to chemical additives, such as chaotropic agents and solvents, and to non-neutral pH (Haki, 2003; de Carvalho, 2011). This is believed to be because the same strategies used by thermophilic organisms to ensure their enzymes continue to function at higher temperatures also provide chemical resistance. It has long been known that there are advantages to performing enzyme reactions in organic solvents, even in (near) anhydrous conditions (Dordick, 1989), and as organic solvent resistance is now more often assayed for, there are a range of enzymes known to be resistant in the presence of a wide variety of different solvents from water miscible alcohols to immiscible solvents like toluene and hexane to halogenated solvents such as chloroform (Doukyu and Ogino, 2010). In addition to increased substrate and product solubility in solvents, advantages include higher rates of reaction, novel enzymatic activities, the ability to recover enzymes via simple filtration in anhydrous solvent conditions, more favourable thermodynamics (preferring synthesis over hydrolysis), enhanced thermostability in (near) anhydrous solvent and suppression of water-dependent side reactions (Ogino and Ishikawa, 2001; Doukyu and Ogino, 2010).

Although it is not yet possible to provide a generalised approach to thermophilicity (Sammond et al., 2016), there are many known features of genomes (Takami et al., 2004b; Frock and Kelly, 2012) and enzymes (Littlechild et al., 2013) which are associated with higher resistance to high temperature. With reference to enzymes, these features can be split into which structural level they affect. At the level of primary structure, thermolabile residues can be replaced with more stable ones; asparagines, glutamines and cysteines, where not catalytically relevant, may be replaced by alanines, prolines, aromatic or charged residues. In terms of secondary structure, flexible loops or linker regions may be shortened or stabilised by interactions with metal ions; α-helices may also be stabilised by capping the helices, i.e. stabilising the dipole that forms from the directional nature of the helix by having negatively-charged residues adjacent to the positively-charged N-terminus of the helix or positively-charged residues adjacent to the negatively charged C-terminus. In the tertiary structure, the structure may be more tightly packed to reduce available space between secondary and supersecondary structures, and increasing stabilisation may be achieved via more ion pairs, hydrogen bonds and disulphide bonds; more hydrophobic residues may also be more appropriately buried in the core of the enzyme, reducing the interactions with bulk solvent and increasing hydrophobic interactions. Lastly, quaternary structure improvements may increase the number of subunits and increase the number of interactions between subunits to further

increase thermostability. It should also be noted that, although this variety of possibilities implies a large difference between mesophilic and thermophilic proteins, in terms of 3D structure there is often little difference and protein sequences may be similar or dissimilar depending on the individual protein in question (Cowan, 1997; Vieille et al., 2001; Sterpone and Melchionna, 2012; Littlechild et al., 2013; Salamanova et al., 2013).

Another substantial advantage of utilising thermophiles is that when processing at higher temperatures, it is less likely that contamination will occur. Common mesophiles will simply be unable to exist at high enough temperatures to cause any issues with a thermophilic process, although obviously thermophiles will remain able to contaminate processes inadequately managed. One example is ethanol production from biomass; when a mesophilic process is used contamination is common, if not constant, reducing yields and profits. Occasionally this may even result in entire failed fermentation runs (Zeldes et al., 2015). At large scale it is also more energy efficient to run biological processes at higher temperatures than 37 ^{o}C. The amount of heat produced by an industrial biological process far exceeds the ability to shed heat, leading to extensive cooling needed to ensure that the process remains stable. Cooling is especially expensive when using temperatures below ambient, as is necessary with mesophilic processes. By using a thermophilic organism, this cooling is either not necessary or greatly reduced (Barnard et al., 2010; Zeldes et al., 2015).

Finally, in a reverse from traditional chemistry, using thermophiles allows for volatile products or by-products to be removed from the system and collected. For instance, when using thermophilic ADHs with isopropyl alcohol or ethanol as sacrificial substrate, it is very easy to drive off the acetone or ethanal produced when operating above their boiling points, favourably shifting the equilibrium of the reaction. Products may also be collected from the bioreactor if they are volatile (Zeldes et al., 2015). This is in contrast to using low boiling-point solvents, typically favoured in bench scale chemistry to allow for the easy purification of products by rotary evaporation of the solvent. In some cases hyperthermophiles may be preferred to drive off more product, and in others thermophiles may be preferred as high temperatures may damage products or impact upon the process. In addition to the previous example of *in situ* product removal, the alternative is an *in situ* subsequent chemical or enzymatic reaction, as both prevent the reverse reaction by sequestering the product (Hilterhaus and Liese, 2009). Continuing the example of thermophilic ADHs, if cofactor regeneration was conducted using a glucose dehydrogenase (GDH), then the gluconolactone formed would hydrolyse into gluconic acid in an aqueous environment. This process is accelerated with increasing temperature and when using water as the solvent, the position of equilibrium would be very far towards gluconic acid, which would positively affect the original ADH catalysed reaction (Goldberg et al., 2007a).

1.4. HISTORY AND TAXONOMY OF *GEOBACILLUS* GENUS

The *Geobacillus* genus was first proposed recently, in 2001 (Nazina et al., 2001). Comprising several species previously described as *Bacillus,* and entirely novel species such as *Geobacillus uzenensis* and *Geobacillus subterraneus*, this genus was formed on the basis of phylogenetic analysis that indicated that this cluster of species formed a group separate from the main body of *Bacillus* species. It should also be noted that *Bacillus* as a genus has been gradually split over the years prior to this claim, with new genera such as *Alicyclobacillus,* and it was the discovery of additional species which necessitated the formation of a new genus to accommodate the genetic heterogeneity in the genus *Bacillus*.

The same work indicated that typical analysis by 16S rRNA sequencing is not adequate for the full delineation at the species level as the sequence similarity could be as high as 97.3%-99.5%

between different species, but could be sufficient to determine the genus of an isolate. Further investigation was carried out based upon genetics, physiology and microbiology of samples, including but not limited to GC content of DNA, DNA-DNA homology, nutritional requirements for growth, cell morphology, and ability of a sample to produce specific acids or other biochemicals.

Subsequent analysis of certain housekeeping genes allows for a more focused annotation of a given isolate, allowing a species to be easily annotated with some confidence. The variance between the different species was incredibly small with 16S rRNA as previously noted, although with other genes such as *recN*, there is a similarity of only 84% at the subspecies level with *G. thermoglucosidasius,* 57% similarity at the genus level and a mere 15% at the phylum level (Zeigler, 2005). Similarly, the genes *gyrB* and *parR* also prove to be useful in this regard (Tourova et al., 2010). This wide margin of difference allows the easy identification of different species within the genus of *Geobacillus*, and should be used in conjunction with existing 16S rRNA methods.

The *Geobacillus* genus contains species that are moderately thermophilic, with typical growing temperatures between 55 ^{o}C and 70 ^{o}C. There has been a gradually increasing level of scientific interest in these organisms in recent years, partly due to their thermophilic nature, and partly due to their ubiquity. *Geobacillus* species have been noted in the Mariana Trench (Takami et al., 2004a) at a depth of nearly 11km. They have also been noted around the world in the soil, even in areas of the world that could not support their obligately thermophilic nature, and this has been a focus of investigation (Zeigler, 2014). The answer for this is believed to be the longevity of spores of this genus, and in particular their size, which allow the spores to be circulated in air currents over very long distances.

Many species from this genus have been investigated for their potential industrial applications. Most obviously, as members of the genus are most often concentrated in areas such as compost heaps and oilfields (Fong et al., 2006; Tourova et al., 2008), they have been thoroughly investigated for their potential in utilising biomass, specifically degradation of lignocellulosic material. One example is the scrutiny for cellulases (Rastogi et al., 2010; Zambare et al., 2011; Assareh et al., 2012), but more typically the focus is on the production of ethanol. In particular *G. thermoglucosidasius* is most frequently investigated for the production of bioethanol in this genus (Fong et al., 2006; Cripps et al., 2009; Tang et al., 2009; Chang and Yao, 2011).

Other uses for this genus have also been proposed, including biosynthesis of silver and gold nanoparticles (Mohammed Fayaz et al., 2011; Girilal et al., 2013), suggested to be of benefit to PCR efficiency, and the production of antibiotics (Ren et al., 2010). Also at organism level, heat and related stress responses have been investigated (Shih and Pan, 2011). Given that this genus is thermophilic, it may be expected that there would be a reduced heat shock response in constituent species; however, a rapid change from 45 ^{o}C to 55 ^{o}C did result in a notable heat shock response in *G. thermoglucosidasius*, regardless of both temperatures being well within the normal growth range of this genus (Tripathy and Maiti, 2014).

A variety of enzymes are also under investigation from this genus, including nitrile hydratases (Chiyanzu et al., 2010), lactonases (Hawwa et al., 2009; Mandrich et al., 2010), lipases (Hutchins et al., 2004), enzymes relevant to alkane degradation (Tourova et al., 2008), acrylamidases (Cha and Chambliss, 2013), and esterases (Yildirim et al., 2009).

Given the aforementioned interest in the *Geobacillus* genus for the production of biofuels and the overlap with other industrial fields, alcohol dehydrogenases are among the most frequently studied enzymes in this genus. Examples have been identified from several species

including *G. denitrificans* (Liu et al., 2009; Ji et al., 2013), *G. stearothermophilus* (Guagliardi et al., 1996; Ceccarelli et al., 2004; Pennacchio et al., 2013a) and latterly *G. thermoglucosidasius* (Lin et al., 2014). Despite this interest, however, investigation into alcohol dehydrogenases in this genus, and specifically into *G. thermoglucosidasius,* remains incomplete. Thus far the best analysis is by Lin et al. (2014) who considered several ADHs with the explicit intention to utilise them for the production of isobutanol from isobutanal and not to survey the enzymes for their wider substrate scope and industrial implications. In addition, the ADHE equivalent in *G. thermoglucosidasius* has been studied by Extance (2012), with the naming convention referring to the *E. coli* bifunctional ADH. Lastly, a further ADH, named ADH-I has also been isolated and studied by Jeon et al. (2008), activity being noted with primary alcohols up to 1-octanol, and 2-propanol.

Moving forward, due to the importance of ADHs (*vide infra)*, it seems likely that the *Geobacillus* genus will be progressively more investigated for novel ADHs given the multiplicity of them already annotated in genomes. It also is probable that the genus will be used as a model for investigating thermophilicity, and some work has already been carried out in this area (Takami et al., 2004b).

Enabling technology, specifically the development of genetic tools to allow more sophisticated manipulation of *Geobacillus* species, is a growing field; for instance, shuttle vectors (Taylor et al., 2008) have recently been developed. A recent review by Hussein et al. (2015) on the *Geobacillus* genus clearly demonstrates the strength of the genus and its potential for industrial use. Of particular note is the detailed consideration of the "genetic tool kit" now available for use and the exceptionally wide range of industries referred to. It is certain that this genus will continue to develop in usefulness and importance in the near future.

In terms of biocatalysis, a *Geobacillus* nitrile hydratase has been immobilised using various supports in bead format (Chiyanzu et al., 2010). This was done due to the relative instability of the enzyme, particularly inhibition when exposed to high concentrations of substrate or product, which is a relatively common feature of enzymes more broadly. Moderate success was obtained, with an increase in half-life from 55 min to 330 min at 60 ^{o}C, and retention of activity (86%) after eight recycles. Furthermore, substrate inhibition was reduced, with a K_i of 195mM compared to a free enzyme K_i of 101mM. This was achieved at a cost of a deterioration in the kinetic constants from a V_{max} of 48.8 µmol mL^{-1} min^{-1} to 4.5 µmol mL^{-1} min^{-1}, K_m from 10.2mM to 17.3mM and k_{cat} from 37,777 min^{-1} to 3,543 min^{-1}. Thus although limited, this success could indicate *Geobacillus* derived enzymes can be effectively immobilised, greatly increasing the applicability of these enzymes to industry.

1.5. ALCOHOL DEHYDROGENASES

Alcohol Dehydrogenases (EC 1.1.1.1 and 1.1.1.2) are some of the most heavily studied enzymes in existence. They are easy to assay, due to their reliance on either NAD(H) or NADP(H) as a cofactor, the reduced forms of which absorb at 340nm and thus can be detected in a UV-vis spectrophotometer. In addition, the reaction that they catalyse, converting an aldehyde or ketone to the respective alcohol or vice-versa, is a frequently performed reaction and one often used to insert chirality into a molecule, for example in the production of a pharmaceutical intermediate. Finally, they are found in all organisms, irrespective of domain. The most famous are likely to be those discovered in *Saccharomyces cerevisiae*, or Horse Liver ADH (HL-ADH), but ADHs in general hold important roles in redox balance and degradation of

aldehydes into less toxic alcohols in species ranging from humans to prokaryotes. They are also involved in biodegradation of alkanes (Ji et al., 2013).

1.5.1. CLASSIFICATION

There are many ways to classify alcohol dehydrogenases. However, four main classes of ADH will be concentrated upon in this work: Short Chain Dehydrogenase/Reductases (SDR), Medium Chain Dehydrogenases/Reductases (MDR), iron containing ADHs and Aldo-Keto reductases (AKRs). Whilst it should be noted that AKRs are not, strictly speaking, ADHs, they may perform the same function and will be considered alongside the more traditional ADHs.

Unhelpfully, these classes are less similar to each other, in terms of sequence identity, than they are to other enzyme classes, which makes identification of an ADH quite challenging without substantial analysis and making automatic annotation risky to rely upon as it is easily possible that an enzyme annotated as an ADH could in fact be a different enzyme class altogether. It is equally possible that the reverse may occur. Another compounding feature is that an enzyme may act as an ADH, but due to limited substrate scope or structure variation, may be reclassified as, for example, a cinnamyl-alcohol dehydrogenase (1.1.1.195), isopropanol dehydrogenase (1.1.1.80) or methylglyoxal reductase (1.1.1.283) (Ying and Ma, 2011). Finally, even ADHs assigned to the same class may exhibit low sequence similarity (sequence identities can be as low as 20%), although conserved ADHs within the same genus are typically similar (Radianingtyas and Wright, 2003). Thus, for the purposes of this work, the phrase alcohol dehydrogenase will refer to any enzyme interconverting aldehydes, ketones and alcohols with the use of cofactors NAD(H) or NADP(H), and may not be solely restricted to the enzyme classes 1.1.1.1 or 1.1.1.2.

1.5.1.1. SHORT CHAIN DEHYDROGENASE SUPERFAMILY (SDR)

These ADHs typically have approximately 250 residues and do not require metal ions for activity or structure. Generally, although not exclusively, these enzymes are associated with smaller substrates. In keeping with the majority of the SDR family, the ADHs typically have only one domain and are involved in essential functions, noted due to the ancient nature of the SDR superfamily, which is found in all organisms. The active site is highly conserved, consisting of either a triad or tetrad of residues: Tyr, Lys, Ser, (Asn).

The SDR superfamily can be split into two main groups, classical and extended, of which extended SDRs tend to be more consistent due to their high chance of being involved in essential functions, whereas the classical SDRs are more varied in their diversity and function (Kallberg et al., 2010). Archetypally these enzymes are dimeric or tetrameric.

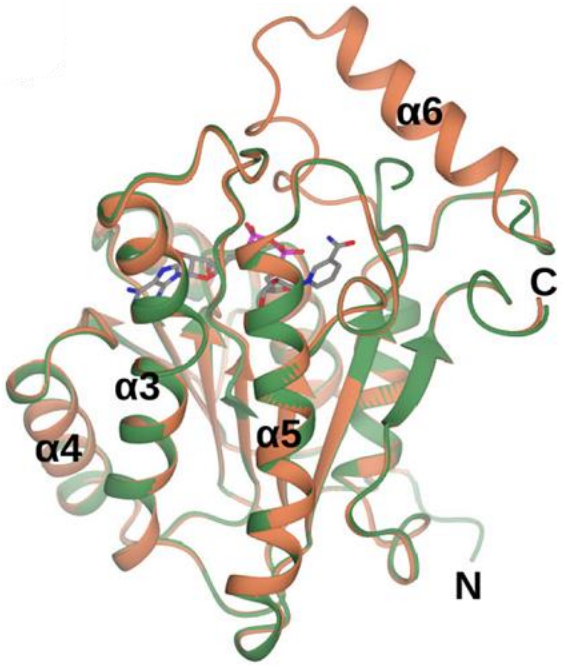


Figure 2 - Monomer of RasADH. The apoenzyme is coloured green, and the holo-enzyme coloured coral. The holo-enzyme is complexed with NADPH, included in cylinder format with carbon atoms shown in grey. Directly reproduced from (Man et al., 2014b).

One particular example of an SDR is an ADH from *Ralstonia* sp. DSM 6428 (RasADH), noted for its ability to catalyse "bulky-bulky" substrates, i.e. those with large hydrophobic groups either side of the ketone (Kulig et al., 2013; Lerchner et al., 2013; Man et al., 2014b). Comprising 249 amino acids, RasADH is a homotetramer with the ability to reduce a wide range of complex aldehydes and ketones including hydroxyketones, ketoesters and exclusively uses NADPH as cofactor (Man et al., 2014b).

Each monomer consists of a single Rossman fold-type domain with a seven-stranded parallel β-sheet flanked by three α-helices on each side. As shown in Figure 2, the addition of the cofactor was required to reveal the last helix, identified as α6 in the figure. It is believed that this helix has a key role by acting as a "lid" to the active site once the cofactor has bound. The active site itself is a hydrophobic tunnel, shown in Figure 3, with the nicotinamide ring of NADPH sitting at the base of the tunnel. Tyr150 is the presumed proton donor for the reductive ADH-catalysed reaction (Man et al., 2014b), and the other amino acids of the catalytic tetrad are noted as Asn112, Ser137 and Lys154 (Lerchner et al., 2013).

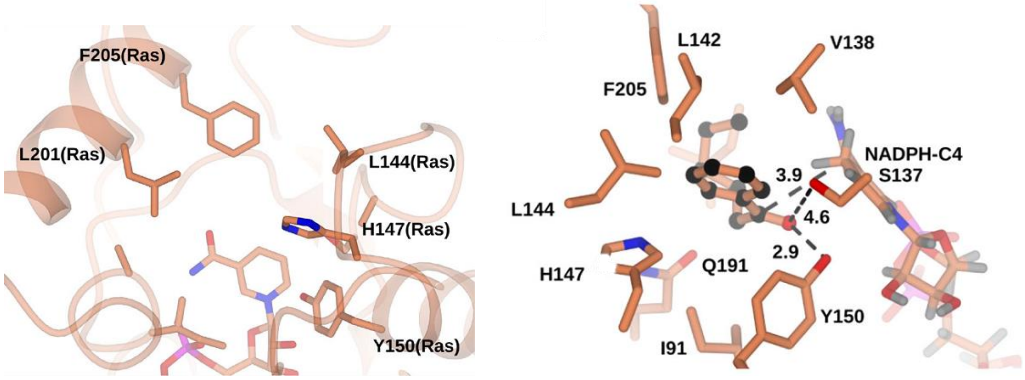


Figure 3 – (Left) Active site of RasADH in ribbon format, illustrating side-chain components and hydrophobic tunnel. (Right) Active site of RasADH with *n*-pentyl phenyl ketone modelled using AUTODOCK VINA with distances given in angstroms. Directly reproduced from (Man et al., 2014b).

With the knowledge that RasADH reduces *n*-pentyl phenyl ketone into the *(S)*-alcohol with excellent enantiomeric excess, it is possible to say that only the *re*-face of the ketone would be presented to the C4-nicotinamide, which delivers the hydride. In addition, the carbonyl group must be within H-bonding distance of the catalytic tetrad, in particular Ser137 and Tyr150. Given this information, the substrate was added to the model, as shown in Figure 3 (Man et al., 2014b). This indicated that Phe205, part of the α6 "lid" helix, was a key amino acid in determining substrate scope due to steric repulsion. Although the chosen substrate was able to fit inside the active site, an aromatic group instead of the *n*-pentyl group would not be able to fit. This is consistent with substrate specificity studies previously conducted (Kulig et al., 2013).

Additional work has been done with this enzyme, specifically to alter the cofactor specificity from NADPH to NADH. A key point is that SDRs typically have either an arginine residue or an aspartic acid residue at the C-terminus of the second β-strand, with the former giving NADPH preference and the latter a preference for NADH (Lerchner et al., 2013). RasADH does not have an aspartic acid residue in the 37th position, and it does have an arginine residue in position 38, consistent with the above pattern. The arginine residue is believed to form a salt bridge with the 2'-phosphate group of $NADP^+$, and hydrogen bonds from other nearby residues such as Asn15 are also instrumental in assisting with the cofactor specificity. With the aim of destabilising these interactions, appropriate point mutations were introduced based upon similar enzymes with preference for NAD^+ (Lerchner et al., 2013).

Using *(S)*-1-phenylethanol as substrate, a RasADH mutant (N15G/G37D/R38V/R39S/A86N/S88A) reduced the K_M for NAD^+ from 3.1 to 0.8mM whilst reducing activity with $NADP^+$ below detectible limits. This gain was somewhat mitigated by the very large drop in k_{cat}, which dropped from the wild-type value of 99 min^{-1} with $NADP^+$ to the mutant value of 1.8 min^{-1} with NAD^+. Similarly, when considering the substrate, the k_{cat} dropped from 47 min^{-1} to 1.38 min^{-1} (Lerchner et al., 2013).

Crystal structures can be found in the PDB with references 4BMN for the apo-enzyme and 4BMS for the holo-enzyme.

1.5.1.2. MEDIUM CHAIN DEHYDROGENASE SUPERFAMILY (MDR)

This is a more complex family, best described by Hedlund et al. (2010): "*The MDRs form a protein superfamily whose size and complexity defeats traditional means of subclassification*".

The superfamily as a whole includes many other enzyme families, which will not be discussed in great detail, but include quinone reductases, cinnamyl alcohol dehydrogenases, threonine dehydrogenases and polyol dehydrogenases, among others. Enzymes of this superfamily may be split into two major groups; one requires zinc ions for activity (Zn^{2+}) and the other does not. Enzymes requiring zinc utilise NAD(H) as cofactor and tend to be dehydrogenases *in vivo*, and those with no need for the metal ion will prefer NADP(H) and tend to be reductases *in vivo* (Hedlund et al., 2010). Industrially speaking, NAD(H)-using enzymes would be preferred due to the relative expense of NADP(H). However, enzymes with a substrate scope encompassing large and complex substrates more useful to industry tend to prefer NADP(H), which is a general problem of ADHs. A typical MDR protein consists of two domains, with the C-terminal domain binding the cofactor and an N-terminal domain binding the substrate (Knoll and Pleiss, 2008; Persson et al., 2008).

The ADHs in the MDR superfamily are the most heavily studied family of ADHs, largely due to the higher number of ADHs (Sulzenbacher et al., 2004). It is also the family most likely to be industrially interesting due to the frequency of larger substrates admitted and the relative propensity of enzymes to accept NADH as a cofactor natively. That said, whilst all ADHs are currently classed as zinc-containing and therefore NAD^+ using dehydrogenases, this is not based on empirical evidence but sequence alignment and prediction. Further, even in the model, there are multiple groupings of ADHs, typically named after their domain, e.g. bacterial ADHs, or named after reactions, e.g. 2,3-butanediol dehydrogenase (Knoll and Pleiss, 2008; Hedlund et al., 2010). The variant groupings are not particularly surprising considering it is believed that up to six evolutionary events occurred to produce MDR enzymes from SDR equivalents, latterly introducing zinc to the enzyme class. Thus it is only to be expected that there is a distinct lack of clarity as to the exact delineation between enzyme families (Jörnvall et al., 2013).

An example of an MDR family ADH is TADH from *Thermus* sp. ATN1, a thermophilic enzyme that utilises NADH to reduce a variety of aldehyde and ketones including 2-pentanone and phenylacetone. It has also demonstrated dismutase ability, i.e. the ability to simultaneously catalyse one aldehyde to carboxylic acid and a second molecule to the corresponding alcohol (Höllrigl et al., 2008). TADH was confirmed to be a 347 residue homotetramer, with each monomer consisting of two domains, as shown in Figure 4. The catalytic domain consists of a core β-sheet of 10 strands, surrounded by five α-helices. The nucleotide-binding domain is a classical Rossman fold domain comprising a central β-sheet of six strands, with six α-helices surrounding it (Man et al., 2014a).

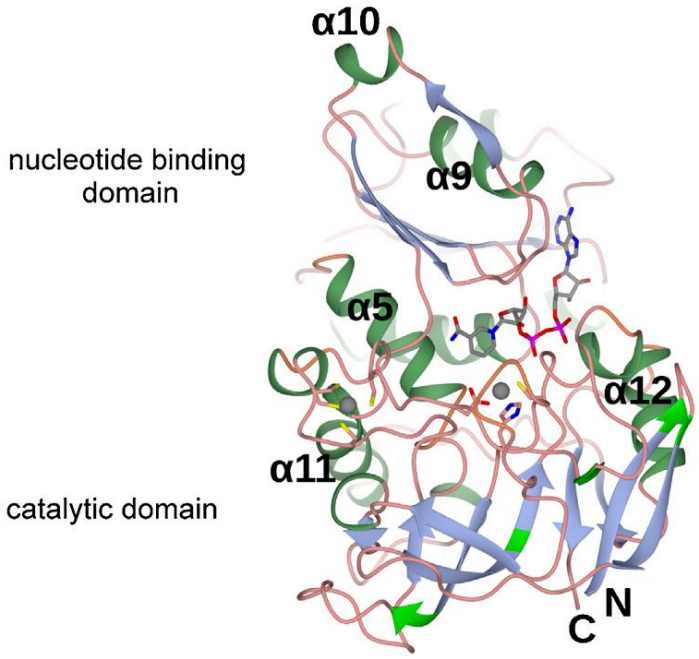


Figure 4 – Monomer A of TADH shown in ribbon format, with NADH and co-ordinating cysteine side chains represented in cylinder format. Carbon atoms are grey and both structural and catalytic zinc atoms are represented as grey spheres. Directly reproduced from (Man et al., 2014a).

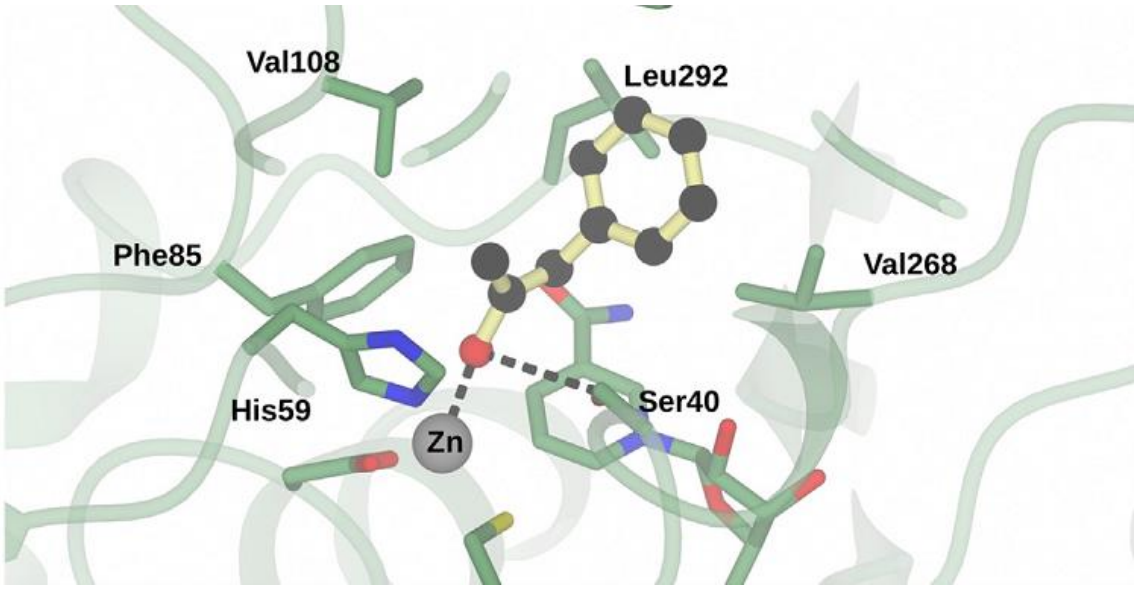


Figure 5 - Model of phenylacetone in the active site of TADH, created using AUTODOCK-VINA. Phenylacetone is shown in ball and stick format, and dashed black lines indicate possible interactions between the carbonyl group of the substrate and both the side-chain of Ser40, a putative proton donor, and the catalytic zinc atom. Directly reproduced from (Man et al., 2014a).

Figure 5 illustrates the active side of TADH. Of particular note is the placement of the C4 nicotinamide carbon, which is approximately 4.5 Å from the side-chain oxygen of Ser40, the putative proton donor in reductive catalysis. The placement of the nicotinamide ring is such that it presents its pro-*(R)* hydride to the active site, matching experimental evidence of *(S)*-selectivity of TADH. Additionally, the active site is largely hydrophobic, which assists in the binding of similarly hydrophobic substrates (Man et al., 2014a).

In comparing this thermophilic ADH with the NADPH-dependent ADH from *Thermoanaerobium brockii* (*Tb*ADH), it shares some features, such as Pro275, which is a conserved residue at the centre of the dimer interface and has been shown to confer thermostability. TADH has 25 proline resides to *Tb*ADH's 21, and *Tb*ADH has only 8 salt bridges as part of the dimer interface, compared with TADH's 14. Increased subunit interactions have been proposed as contributing to the increased solvent tolerance of enzymes (Man et al., 2014a).

As might be expected, there are differences in the cofactor binding loop, with *Tb*ADH securing the 2' ribose phosphate of NADPH through interactions with Arg200, Tyr218 and Ser199. TADH excludes NADP through the negatively charged Asp200 and hydrophobic Ile202, which replaces Ser199 of *Tb*ADH. Lastly, the orientation of the nicotinamide ring in the substrate pocket is different, as can be seen in Figure 6. *Tb*ADH has the ring rotated approximately 60° down towards the side-chain of Trp110, substantially altering the active site and in particular exposing the *pro-(S)* hydride of the C4 carbon of the ring to the substrate-binding pocket. It is believed that this change may be the reason for small substrates being converted with inverted stereoselectivity with *Tb*ADH compared to larger substrates, whereas TADH converts all substrates to *(S)*-alcohols.

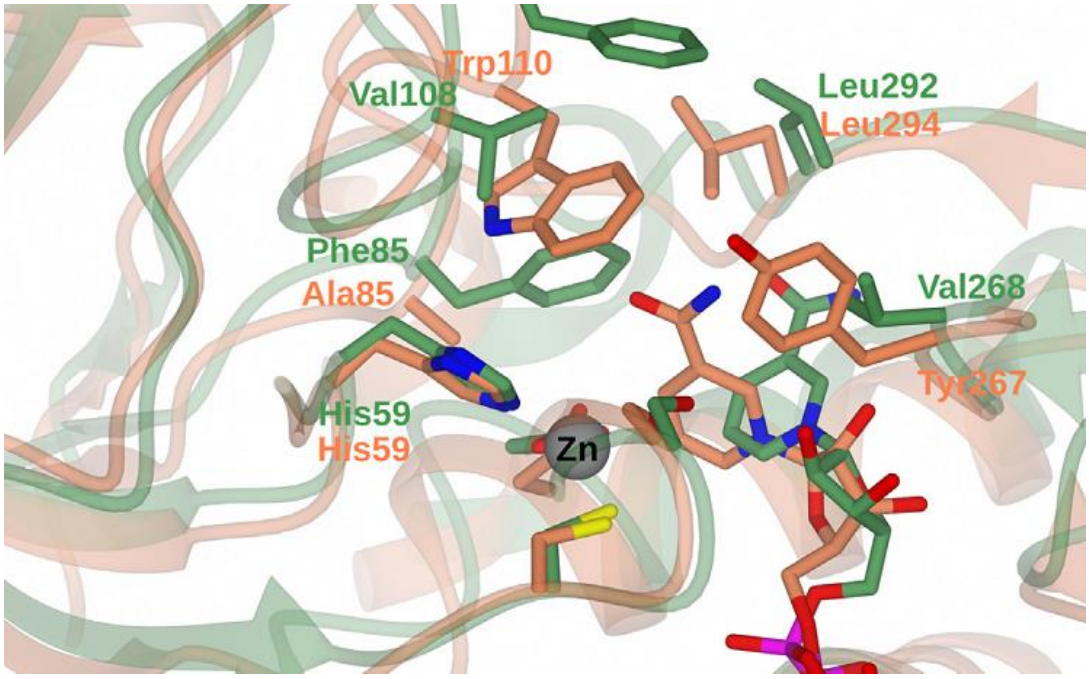


Figure 6 – Superimposition of active sites of TADH and TbADH, with residue side-chains shown in cylinder format in green and coral colouring respectively. Directly reproduced from (Man et al., 2014a).

The PDB record for TADH is 4CPD.

1.5.1.3. IRON CONTAINING ADHS

Despite the naming convention, these ADHs may use either zinc or iron ions for activity, and may in fact use both. There is a much larger range of size for these enzymes, noted from approximately 385 residues to over 900. These enzymes are typically of greater potential than any other family in terms of their substrate scope; however, ADHs requiring iron for activity commonly have weakly-bound ions, meaning that the ion can dissociate during purification.

Some of these enzymes may regain activity with the subsequent addition of iron ions to their buffer, others may not (Tamarit et al., 1997). In the case of those enzymes that can have their activity restored, one approach may to be replace Fe^{2+} ions with Co^{2+} for purification and then replace with Fe^{2+} again to restore higher levels of activity (Tse et al., 1989).

If this were not enough, iron containing ADHs are also often oxygen sensitive, highly sensitive in many cases, making purification and testing incredibly difficult if not done in the complete absence of oxygen. Some enzymes can regain activity over time if stored in the presence of iron (Tamarit et al., 1997), or strictly anaerobically, but this is not true in all cases. Finally, these enzymes are more likely to utilise NADP(H) as cofactor rather than NAD(H), which may be an issue for industrial purposes (Radianingtyas and Wright, 2003).

A particularly well studied example of an iron-containing ADH is ADH2 from *Zymomonas mobilis* (Neale et al., 1986; Mackenzie et al., 1989; Tse et al., 1989; Tamarit et al., 1997; Moon et al., 2011). It is a dimeric, NADH-dependent, ADH with 383 residues.

A monomer of zmADH2 forms 17 α-helices and 9 β-strands and has two domains, as shown in Figure 7. The first domain forms a globular structure with α/β mixed folding, comprising seven β-strands in one β-sheet and two α-helical bundles containing four and two α-helices. The C-terminal domain, also globular in structure, is entirely composed of α-helices in two bundles, one of three α-helices and another of eight α-helices. The C-terminal domain is where the iron ion binds, and the cofactor binds in the cleft between the two domains (Moon et al., 2011).

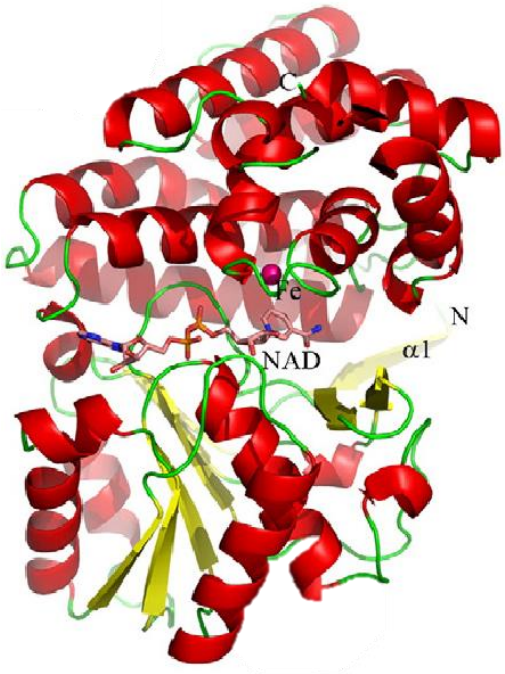


Figure 7 - Monomer of zmADH2 in ribbon form, complexed with iron ion and NAD^+ cofactor. The iron ion is shown as a pink sphere and the bound NAD^+ as a stick model. Directly reproduced from (Moon et al., 2011).

Compared to previous ADH families there are two notable differences. The first is that the nicotinamide ring is ambiguously placed, which implies flexibility in this area of the enzyme, and could further imply a lack of stereocontrol. The second is that the cofactor binding site is located within a hydrophilic crevice, rather than hydrophobic as has been the case in the previous two examples. Following from the hydrophilic nature of the active site, which is exposed to bulk solvent, a water molecule co-ordinates with the iron ion, as shown in Figure 8. The last key point with regards to the cofactor is the presence of Asp39, which is not

conserved in all similar enzymes. Asp39 appears to be critical for ensuring NAD^+ specificity rather than $NADP^+$ specificity, and when it is replaced by glycine in enzymes such as 1,3-propanediol oxidoreductase, then the specificity switches to $NADP^+$ (Moon et al., 2011).

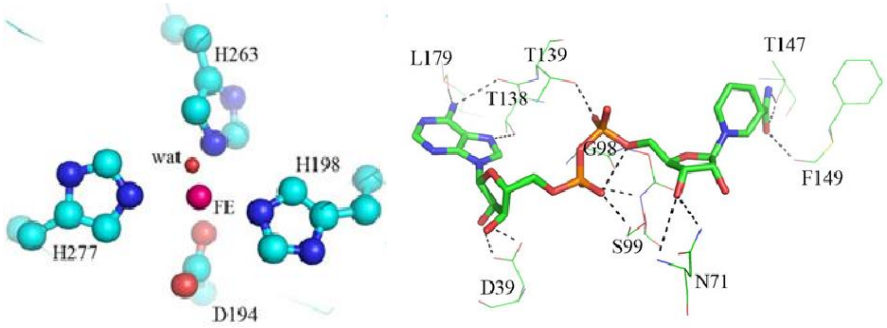


Figure 8 – (Left) Binding site of the iron metal (pink sphere) in zmADH2. Note the water molecule (red sphere) co-ordinating to the iron. Carbon is coloured cyan, nitrogen blue and oxygen red; all ball and stick models. (Right) Schematic representation of NAD^+ interaction in the binding pocket, formed along the interface of the two domains of the enzyme. Residues are represented as line models, the NAD^+ as a thick stick model and hydrogen bonds formed between the two are dotted black lines. Directly reproduced from (Moon et al., 2011)

Overall, the active site is likely to be constrained by the hydrophobic residues Phe254 and Phe149, in addition to His267. That said, the modelled ethanol is >3.2Å from His267, which is greater than the hydrogen bond distance, implying a lack of substantial direct interaction. A water molecule could act as mediator, and this has been noted elsewhere. These residues are noted in Figure 9 along with other key residues in the active site (Moon et al., 2011).

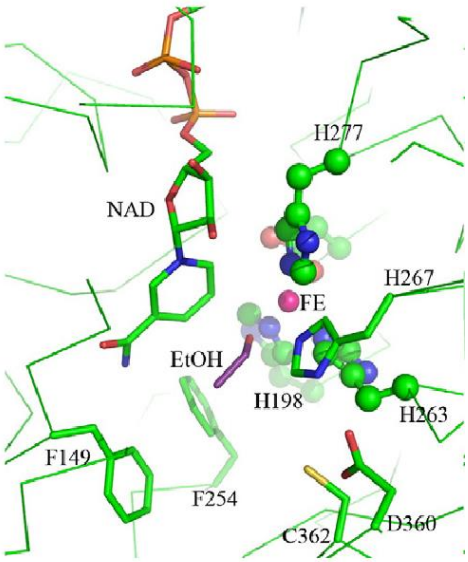


Figure 9 - Modelling of ethanol as substrate in the active site of zmADH2. All relevant residues are indicated as stick models and labelled. EtOH is also presented as a stick model, with carbon atoms in violet and oxygen in red. Directly reproduced from (Moon et al., 2011).

Of additional interest is that bifunctional enzymes, where acetyl-coA is catalysed to ethanol, most often have a member of the iron-containing ADH family as their ADH domain. As the ADH from *G. thermoglucosidasius* has been crystallised (Extance et al., 2013), this is a convenient example to use.

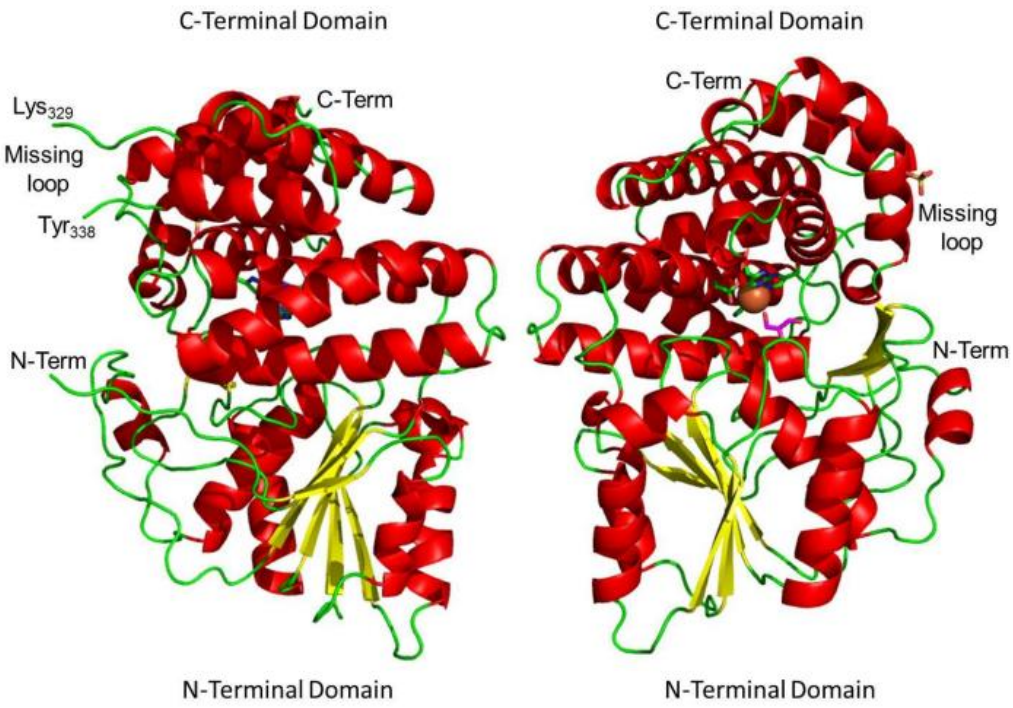


Figure 10 - Structure of the ADH domain from the bifunctional *G. thermoglucosidasius* ADHE with front and back views. The zinc ion is represented by a brown sphere and glycerol is coloured. The missing loop refers to a small section that is too mobile to be accurately determined. Directly reproduced from Extance (2012).

There appears to be a great similarity between standard iron-containing ADHs and the ADH section of ADHE, at least in tertiary and quaternary structure. Comparing Figure 8 and Figure 10, it can be seen that both have two domains, one of which is entirely composed of α-helices, and the other being a Rossmann-fold type domain, with α-helices either side of a β-pleated sheet. Whilst the latter enzyme contains zinc rather than iron, as mentioned previously, the so-called iron-containing alcohol dehydrogenases are able to utilise zinc as cofactor, and may retain significant activity with other cofactors as well. One of the differences is with regard to the "missing loop", which is only conserved in ADHs as part of bifunctional enzymes, and not stand alone enzymes; this is taken to mean that it is involved in the co-ordination between the AldDH and ADH domains in the bifunctional enzyme and therefore would not be found elsewhere (Extance et al., 2013).

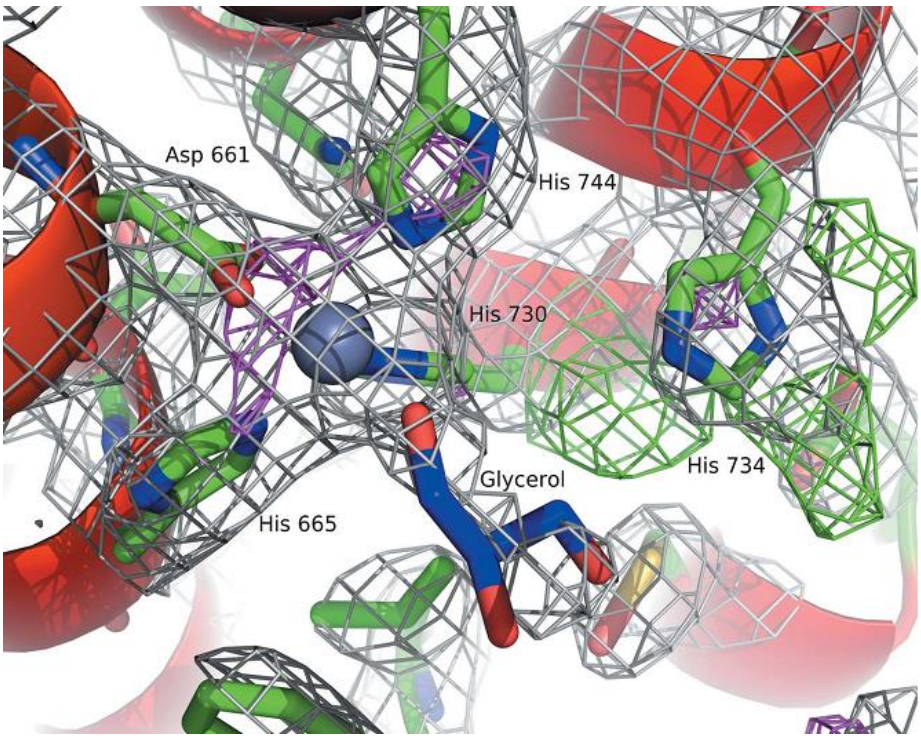


Figure 11 - Metal ion binding site of the ADH domain of the *G. thermoglucosidasius* ADHE, overlaid with electron density. Co-ordinating residues are shown in stick form with green carbon atoms; zinc ion is represented as a grey sphere and glycerol with blue carbon atoms. α-Helices are shown as red spirals, grey mesh represents the electron density from the final $2F_o$-F_c map contoured at 1σ, green mesh represents the positive difference (F_o-F_c) density and magenta mesh the negative difference density, contoured at +3σ and -3σ respectively. Directly reproduced from Extance et al. (2013) with permission of the International Union of Crystallography. See http://journals.iucr.org/ for further information.

Due to the increase in activity when additional metal ions are provided, it is believed that the zinc ion is not merely structural, but catalytic, and is believed to polarise the carbonyl of the substrate prior to NADH hydride transfer. The metal ion is octahedrally co-ordinated with Asp661, His665, His730, His744 and a glycerol molecule, as shown in Figure 11. All residues are part of the α-helical domain (Extance et al., 2013).

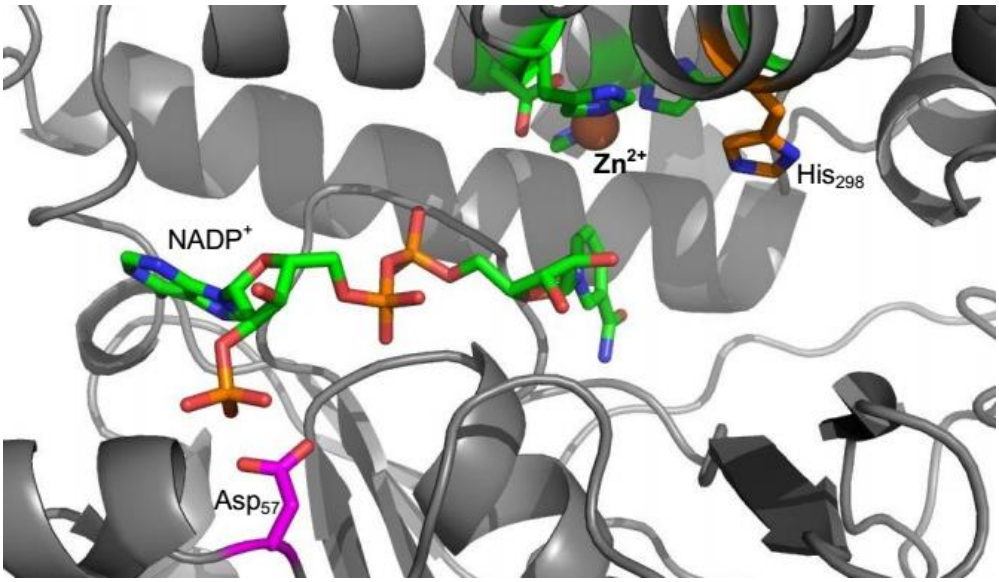


Figure 12 - Depiction of $NADP^+$ superimposed into the *G. thermoglucosidasius* ADH crystal structure of ADHE. Magenta residue is Asp57 and the orange residue is His298. Directly reproduced from (Extance, 2012).

Of particular interest to cofactor specificity is His298, which has been found in similar enzymes. The His298 equivalent allows for NADH specificity; however, when mutated to Arginine, specificity alters substantially towards NADPH. That said, overall activity was drastically reduced as well, potentially due to interference with the metal ion (see Figure 12). There are also ambiguous signs that Asp57 may be a key residue as well, yet the significant distance between this residue and the NADPH phosphate group (18.4Å) would appear to preclude such possibilities (Extance, 2012).

PDB references for zmADH2 are 3OWO (without cofactor), 3OX4 (with cofactor), and reference for the ADH section of ADHE from *G. thermoglucosidasius* is 3ZRD.

1.5.1.4. ALDO-KETO REDUCTASES SUPERFAMILY

This superfamily contains monomeric enzymes of approximately 320 residues with either exclusive preference for NADP(H) as a cofactor, or will accept NAD(H) whilst preferring NADP(H). In rarer cases NADH preference can be found, but almost none are exclusively NADH dependent. As with the previous superfamilies, the AKRs include a wide range of enzyme families, including hydroxysteroid dehydrogenases, mannose reductases and potassium channel β-subunits. Another key difference between AKRs and other types of ADHs are that they do not adopt a classic Rossmann fold, rather a TIM or $(\beta/\alpha)_8$ barrel. AKRs are also more often monomeric, but may be dimeric, trimeric, tetrameric or octameric (Di Luccio et al., 2006; Lapthorn et al., 2013).

Examples of AKRs are YtbE and YvgN, both from *Bacillus subtilis*. The former is considered an essential protein in *B. subtilis* for the detoxification of aldehydes (Commichau et al., 2013) and the latter has been implicated in formulation of spores (Lei et al., 2009). Due to their high sequence identity (70%), they will be considered as a pair for the purposes of this example.

AKRs have a typical catalytic tetrad of tyrosine, histidine, aspartate and lysine, which is essential for reactions to take place. In addition, loops β1 and β7 directly interact with each other via hydrogen bonding, which forms a "safety belt" to secure the pyrophosphate group of the cofactor; it is believed that a conformational change to release $NADP^+$ from this safety belt is the rate-limiting step in reactions. In the case of YvgN, the safety belt is stabilised by a hydrogen bond between Lys26 and Gln194, which does not induce a significant conformational change. YtbE replaces Lys26 with Gln26 and has no apparent safety belt (Lei et al., 2009).

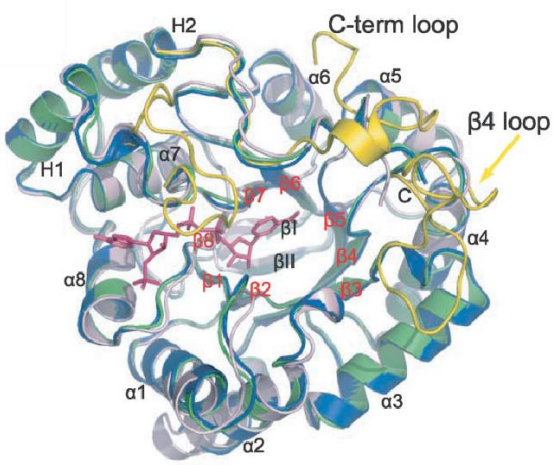


Figure 13 - Ribbon diagrams of apo YtbE, apo YvgN and holo YvgN coloured in salmon, blue and green respectively. The bound NADPH in holo YvgN is represented by sticks, and the yellow areas refer to FR-1 from *Mus musculus*, a typical eukaryotic AKR. Directly reproduced from (Lei et al., 2009)

Particularly interesting about the structure of these AKRs from *B. subtilis* is the substantial difference when compared to traditional eukaryotic AKRs. This difference is most easily seen in loops β4 and β7, and these define the substrate binding pocket and the cofactor binding pocket, respectively (Lei et al., 2009). These differences are highlighted in Figure 13 above.

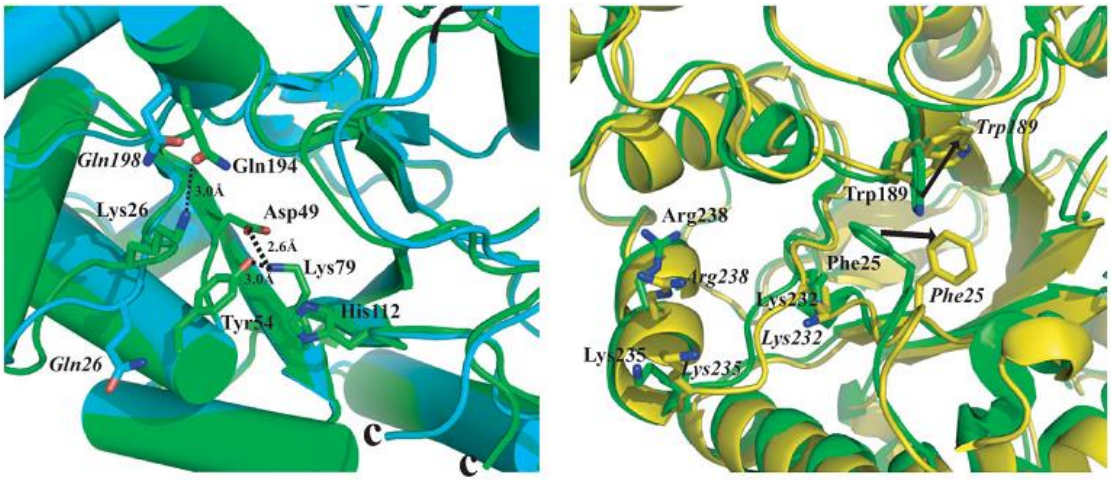


Figure 14 - (Left) Structures of YvgN (green) and YtbE (cyan) showing the "safety belt" (dashed line) and catalytic tetrad (Tyr54,Lys79, Asp49, His112). (Right) Superposition of apo YvgN (green) and holo YvgN (yellow).

The catalytic tetrad is Tyr54, Lys79, Asp49 and His112 for YvgN, and Tyr54, Lys83, Asp49 and His116 for YtbE. The former tetrad is shown in Figure 14. It is believed that a conserved hydrogen-bonding network between the catalytic tetrad (as indicated) depresses the pKa of tyrosine to enable it to act as an acid to facilitate the hydrogen transfer to the substrate. The His112 residue is buried in a hydrophobic environment consisting of Ile53, Trp81 and Trp113 and is therefore believed to be responsible for the orientation of the carbonyl in the active site via steric hindrance (Lei et al., 2009).

No substantial shift in secondary structure was noted upon NADPH binding to YvgN although Lys232, Lys235 and Arg238 reorientated to form hydrogen bonds with the 2'-phosphate of the cofactor. Trp189 and Phe25 also moved to allow space for the cofactor to bind. The cofactor

itself was tightly stabilised by a hydrogen bonding network, the previously mentioned safety belt and aromatic stacking interaction with Trp189 (Lei et al., 2009).

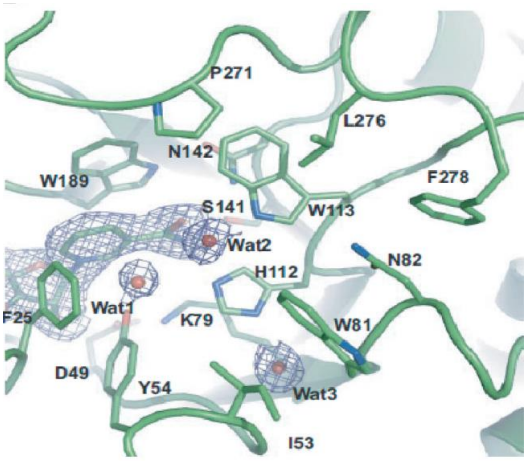


Figure 15 – Putative substrate binding site of YvgN, which includes three water molecules as red spheres. The blue mesh represents electron density with the $2F_o$-F_c density map (σ=1.0). Directly reproduced from (Lei et al., 2009).

The substrate binding pocket in YvgN is believed to be mainly composed of aromatic and apolar residues Phe25, Ile53, Tyr54, Lys79, Trp113, Trp81, His112, Asn82 and hydrophobic residues found in the C-terminal loop: Pro269, Leu274, Phe276. Three water molecules are also present in the binding pocket, which are believed to form hydrogen bonds with Tyr54, Trp113 and His112, which assist in orientating the carbonyl groups of the substrates during reactions (Lei et al., 2009).

Despite a wide range of accepted substrates, both YvgN and YtbE showed remarkable activity with heteroaromatic aldehydes such as 4-nitrobenzaldehyde. Replacing the 4-nitro group with a 4-chloro or 3-nitro group substantially reduced the activity of both enzymes. Substituting with a 2-halogen instead resulted in substantial increases in activity. From human AKRs it has been previously shown that Arg311, positioned in the C-terminal loop, determines the substrate specificity via hydrogen bonds with negatively-charged groups on the substrate, yet neither YvgN nor YtbE have an equivalent residue. In common with prokaryotic AKRs, Asn 82(YvgN)/Asn86 (YtbE) is highly conserved and may have a similar role in determining substrate specificity through interaction with the 4-group of the aromatic ring. YvgN and YtbE were similar in substrate scope and structure in this regard (Lei et al., 2009).

It has also been suggested that AKRs may act as dehydrogenases, i.e. by converting alcohols to their respective aldehydes and aldoses, but no activity was noted with a range of substrates and $NADP^+$ as cofactor (Lei et al., 2009).

Further work has been conducted with YtbE, due to its industrially-interesting substrate scope (Ni et al., 2011). The enzyme was investigated for its ability to reduce methyl *o*-chlorobenzyolformate to methyl *(R)-o*-chloromendelate, a key intermediate in the production of *(S)*-clopidogrel, a treatment for atherosclerosis. The issue was stability of the enzyme in industrially relevant conditions, including thermostability at temperatures above 45 °C, insufficiently low activity at 20 °C for an economically effective industrial process, substrate inhibition, and product inhibition. A particular problem was product inhibition, which led to permanent inactivation of the enzyme through unfolding. Several additives were tested to determine if they could increase the stability of YtbE, including glycerol, glucose and other low-

molecular weight molecules which previously have shown to strengthen hydrophobic interactions among non-polar amino acids as a strategy to stabilise proteins against denaturation. Incubating YtbE with 30% glycerol for 30 minutes led to near complete retention of the enzyme activity, best of all additives studied (Xu et al., 2014).

Given the above structure, the active site is mostly composed of non-polar and aromatic residues, which would be amenable to stabilisation through the use of low-molecular weight molecules. The reason why glycerol may be the best stabilising agent could be due to the three water molecules normally found inside the active site. It is plausible that some or all of these were replaced with glycerol, which may provide additional stability through increased size compared to water. It is also possible that the additional alcohol groups can simply allow for additional hydrogen bonds to be formed or that due to the carbon backbone of the molecule, glycerol is able to form hydrogen bonds whilst stabilising itself with more hydrophobic interactions with the existing residues in the active site. A third option is that this is merely another example of glycerol stabilising a protein, as it is a well-known generic stabiliser (Gekko and Timasheff, 1981; Timasheff, 1992).

PDB records for these examples are as follows: apo YtbE (3B3D), apo YvgN (3F7J), holo-enzyme YvgN (3D3F).

1.5.2. ADVANTAGES AND USES OF ADHS

All ADHs, irrespective of their family, catalyse the reaction between alcohol and aldehydes and/or ketones, with the aid of a cofactor, either NAD(H) or NADP(H). This is a crucially important chemical reaction which is frequently performed throughout the chemical industry, in both directions. In areas such as the pharmaceutical industry, alcohols are preferred as products, for this produces a chiral product, which can either be kept as is, or subsequently converted into whichever molecule is required (Ziegelmann-Fjeld et al., 2007). The cosmetic industry will usually prefer specific enantiomers of alcohol too, as these may have specific fragrances which are needed for different purposes. Further, alcohols may be preferred as an intermediate to esters, which tend to be widely used as fragrances (Welsh, 1989).

The bioremediation sector is also likely to prefer alcohols as well, but solely due to their safety. Although alcohols may remain toxic, they are substantially less toxic than their aldehyde counterpart; in areas such as the paper or textile industries, or even wastewater treatment facilities, there may be high concentrations of methanal, ethanal and other small aldehydes that pose a very substantial health risk, compounded by their volatility.

More broadly, there are a variety of alcohols considered useful biofuels, bioplastics or key intermediates for the chemical industry. These include isobutanol, butanol, 1,3-propanediol and 1,4-butanediol (Jarboe, 2011; Erickson et al., 2012). Much work has concentrated on the biosynthesis of small alcohols, which will typically involve an ADH as the last step.

Conversely, there can be occasions where the reverse reaction is preferred, with one notable example being the raspberry ketone. In this instance there is no technical reasons, however due to legal and consumer choice reasons, it is preferred to have products which can be labelled as natural. Therefore the ketone can be directly extracted from raspberries, but as there is only 3.7mg kg^{-1} available, this ensures a great cost. There are several potential routes available, both biological and chemical, however many do not qualify for the legal standard of natural, and therefore cannot be used. One option is to use rhododendrol, the alcohol equivalent, using an ADH to convert the alcohol to the ketone. The rhododendrol can be

extracted as the corresponding 2-glycosides from the bark of silver tree, and is converted into the requisite alcohol by β-glycosidase. As the raw material is directly extracted from a natural source and a purely biological (i.e. enzymatic) process is followed, this meets legal requirements (Welsh, 1989; Edegger et al., 2006; Wenda et al., 2011).

1.5.3. CHALLENGES OF ADHS

There are several major disadvantages of using ADHs, which afflict most, if not all, enzymes. Being adapted to a cellular environment, they are typically adapted to accept a concentration of substrate within a narrow band, usually a very low concentration. Equally there is rarely a large amount of product that remains, as enzymes do not work in isolation, being part of a metabolic pathway. Even if substrates and products were not toxic and naturally inhibitory to both whole cells and isolated enzymes, the conditions required for an accepted industrial process differ wildly from those anticipated in a cell.

Several options have been considered for improving the stability of ADHs in industrially relevant operation conditions. Firstly is mutagenesis, engineering the enzyme itself to be inherently more stable, whether by eliminating substrate inhibition or by otherwise improving resistance to the intended industrial conditions (Ziegelmann-Fjeld et al., 2007). Switching from whole cell to isolated enzyme or the reverse can also assist. If a biosynthetic approach can be obtained, then there is less substrate and product exposed to the enzyme, particularly if an excretion system is employed so that the final product is obtained from the supernatant rather than the cells. Conversely, if isolated enzymes are used, then weaknesses deriving from a whole cell system such as resistance to additives or solvents may be obviated.

A third option is stabilisation of the enzyme. This may take many forms, including immobilisation of the enzyme which may increase the stability of the enzyme, but also allows the enzyme to be recovered with ease, paving the way for reuse of an enzyme (Spickermann et al., 2014). Typically this involves physical adsorption of isolated enzymes (potentially reversible) or covalent bonding (essentially irreversible) to solid supports (Hollmann et al., 2011b). Other ways of immobilisation include binding enzymes to a support, entrapment or encapsulation of the enzyme, and cross-linking (Knezevic-Jugovic et al., 2011).

A wide range of protocols for stabilising and immobilising enzymes is given by Minteer (2011); however, specific options include covalent bonding of reactive amino groups to cyanogen bromide activated agarose beads, multipoint covalent attachment to glyoxyl agarose, cross-linking with glutaraldehyde, and anionic exchange followed by covalent reaction to amino-epoxy supports (Rocha-Martin et al., 2009). Further possibilities include forming cross-linked enzyme aggregates (CLEAs) by precipitating the enzyme before cross-linking the aggregates, potentially with glutaraldehyde as previously mentioned (Sheldon, 2012); linking to amino-functionalised carriers such as porous glass beads, magnetic particles (magnetite) and nanodiamonds, with glutaraldehyde (Goldberg et al., 2008); cross linking with 1-ethyl-3-(diamino-propyl) carbodiimide (EDAC) to commercial support materials such as Eupergit® (Chiyanzu et al., 2010); covalent bonding to amino-epoxy methacrylate resin (Grimaldi et al., 2016); and whole cell encapsulation in alginate (Kurbanoglu et al., 2011; Leuchs and Greiner, 2011) or sodium cellulose sulphate (Goldberg et al., 2007a).

As is implied from the short list above, there are a large number of methods available, but immobilising the enzyme or whole cell is not always a positive move. Stability to reaction conditions, thermostability and recoverability of the enzyme may be increased, but immobilisation may also result in decreased activity and loss of specificity. Methods such as

crosslinking can be non-specific themselves, resulting in changes to the enzyme. There is no overarching strategy that can always be followed, meaning that each enzyme and process has to be considered on a case by case basis.

ADHs also suffer from an additional issue, cofactor regeneration. As they rely upon NAD(P)(H) to mediate the catalysis of their substrate, and the cofactor is used stoichiometrically, then it is necessary to replace the cofactor. In industrial processes this is particularly important both due to the scale of the reaction but also the extreme cost that large quantities of cofactor would attract, especially NADPH which is ten times more expensive than NADH. When in a whole cell situation the problem is slightly different as the host organism can regenerate the necessary cofactor if it has sufficient glucose or other food source. The issue then becomes whether the demands of the biosynthesis overwhelm the host cell's ability to regenerate the relevant cofactor or whether the particular ADH involved has an affinity for the cofactor greater than other enzymes involved in primary metabolism. It is possible for the drain on cofactor to be so substantial that it can interfere with the good running of the cell and cause cell death. The most prominent example of this is furfural, liberated during the processing of lignocellulosic feedstocks, it is highly toxic to most organisms used in the processing, and this is due to ADHs converting the furfural to furfuryl alcohol, rapidly removing all available NADPH from the cell, exhausting all supplies that are necessary for primary metabolism to function (Miller et al., 2010).

There are several methods used to regenerate the cofactor, whether it is NADH or NADPH required. The first is to use a sacrificial second substrate alongside the required substrate. In the case of ADHs, this would typically be either ethanol or isopropanol, small alcohols that can be acquired cheaply. It is even possible to use a small alcohol as a solvent or co-solvent, dissolving the actual substrate and/or product into it. If the ADH in question is resilient to this system, and able to accept both substrates, then this can be an elegant situation. Furthermore, when using thermophilic ADHs this situation improves dramatically as ethanal and acetone are both exceptionally volatile, and therefore are easily removed via an *in situ* recovery system. This ensures that the system is irreversible and will assist in driving the reaction to completion, which is particularly relevant as large molar excesses are often needed of the sacrificial substrate to ensure that the primary reaction occurs (Hollmann et al., 2011b; Wenda et al., 2011).

A second option is to use glucose dehydrogenase (GDH) in a two enzyme approach, which converts glucose into gluconolactone, which irreversibly hydrolyses into gluconic acid in the presence of water. As glucose is cheap and readily available, and as GDH is able to utilise both NAD^+ and $NADP^+$ as cofactors, this is a very useful method. The irreversible second stage ensures that the reaction will eventually run to completion, although care may need to be taken, as with all cofactor regeneration methods, that the regeneration is not hindering the primary reaction by being the rate-determining step. GDHs are well known and characterised; therefore a suitable enzyme can be chosen to work with reaction conditions as required. The major drawback in this case is the pH change which will invariably occur, challenging buffers, particularly at larger scales. Even if the pH change can be controlled, which is possible with suitable engineering control systems such as proportional-integral-derivative (PID) control, the salt produced will affect the ionic strength and the volume of waste will be quite substantial (Huisman et al., 2010).

Similar to the use of a GDH is the use of a formate dehydrogenase (FDH), which catalyses the formation of carbon dioxide from formic acid, which is effectively irreversible due to the ease at which the gaseous product can be emitted. Whilst there still is a pH change, this can be

more easily remedied through continual addition of formate to the reaction. FDHs can reduce both NAD^+ and $NADP^+$ making them useful in both systems (May, 2009; Gao et al., 2014).

Further options for cofactor regeneration do exist, including electrochemical regeneration, photosynthetic based regeneration and alternative enzymes such as lactate dehydrogenase, glutamate dehydrogenase, phosphite dehydrogenase, NADH oxidase (Wu et al., 2012)and hydrogenase. Most other enzymes would have similar problems to both the FDH and GDH previously mentioned, however as NADH oxidase relies on a continuous supply of oxygen gas and the hydrogenase relies upon hydrogen being freely available, both may additionally pose issues with regards to diffusion of the gas throughout a large bioreactor as well as fire or explosion risk. The other examples given are less well established and rarely used at industrial scale due to the relative ease and the effectiveness of the biological approach, and most often ADHs, GDHs, or FDHs (Chenault and Whitesides, 1987; Goldberg et al., 2007a; Moore et al., 2007; Hilterhaus and Liese, 2009; May, 2009; Huisman et al., 2010; Hall and Bommarius, 2011; Hollmann et al., 2011a).

Cofactor regeneration is a particular issue when using isolated thermophilic enzymes because NAD(P)H is not thermostable. Their thermolability is exacerbated by acidic pH, ionic strength and other buffer components such as phosphate and citrate (Wu et al., 1986; Chenault and Whitesides, 1987). The oxidised and reduced forms of the cofactors have different degradation pathways and also different sensitivities; NAD(P)H is more easily degraded in acidic conditions whereas $NAD(P)^+$ is resistant to acidic conditions but easily degraded in alkaline conditions. Given the nature of redox reactions this makes determining suitable reaction conditions extremely difficult (Chenault and Whitesides, 1987). Often cofactor stability is overlooked in favour of regeneration, yet it is clear that even with an effective regeneration system, even at moderate temperatures, cofactors will degrade (Wu et al., 1986; Chenault and Whitesides, 1987; Pennacchio et al., 2011). It is imperative that the cofactor is stabilised to ensure an effective industrial scale system.

Several routes are available. The first is to ensure that there is sufficient enzyme activity and cofactor such that loss of cofactor is irrelevant, something particularly done at bench scale (Pennacchio et al., 2011, 2013b). The second is to use the cofactor in the solid phase, which massively improves its thermostability (Kulishova et al., 2010); another is to design a cofactor regeneration system that is able to regenerate degraded cofactor, and the first steps towards this have been done (Honda et al., 2016), but it is an incredibly complex system necessitating many more enzymes and only works for NAD^+ regeneration thus far. It is also possible to immobilise the cofactor on a solid support, or immobilise the enzymes used for cofactor regeneration on a solid support (Rocha-Martín et al., 2012; Gao et al., 2014). None of these options is perfect currently, but this is a continuing area of investigation; the first option is probably the most likely to be taken, but this drives up the cost and would potentially not allow thermostable enzymes to be used to the best of their ability if a short reaction time is preferred.

1.6. AIMS AND OBJECTIVES OF THE PROJECT

The aim of this project is to identify novel enzymes, specifically alcohol dehydrogenases, that are of potential industrial use. These enzymes will be acquired from *G. thermoglucosidasius* for two principal reasons. Firstly, multiple ADHs are already under investigation in the Centre for Extremophile Research for their role in ethanol production *in vivo* and for the potential production of butanol. These previously isolated enzymes are ideal as a starting point to determine the potential industrial significance of enzymes from this species. Secondly, *G. thermoglucosidasius* is a moderate thermophile that is being investigated to determine whether it is suitable to produce ethanol from municipal waste at industrial scales. As part of previous investigations it has been determined that numerous ADHs exist and are of unknown function and substrate scope. Given the thermophilicity of *G. thermoglucosidasius*, these enzymes are more likely to be thermophilic and solvent tolerant than mesophilic enzymes. These features are in particular demand in industrial processes and were of high importance to the project.

Key objectives for this project are:

- Identify and isolate genes encoding for putative ADHs, followed by cloning into *E. coli*
- Determine the substrate scope of acquired recombinant enzymes
- Analyse the enzyme kinetics with several substrates of relevance to research or industry
- Identify products using Gas Chromatography (GC) or High Perfomance Liquid Chromatography (HPLC) methods, as appropriate
- Determine stereoselectivity of enzymes using chiral GC methods, as appropriate
- Evaluate stability of enzymes in industrially-relevant conditions, including thermostability and solvent stability
- Identify prospective applications for enzymes and demonstrate the potential for using them in these industrial or research applications by scaling up as necessary to determine proof of concept reactions based upon existing or novel reactions

By achieving these key objectives, it will be possible to ascertain whether *G. thermoglucosidasius* can be a useful source of enzymes for further biotechnological applications and inform further research projects involved with this species.

Initial results were obtained from Alcohol Dehydrogenases already under investigation. They were all obtained from other members of the Centre for Extremophile Research who were examining them for their potential in the biosynthesis of ethanol and higher alcohols as biofuels. All genes were obtained from *G. thermoglucosidasius* and expressed in *E. coli,* provided as glycerol stocks of the production strain BL21 (DE3), using the plasmid pET28a to provide an N-terminal hexa-histidine tag for purification purposes.

2.1. ADH A

The first Alcohol Dehydrogenase (ADH) studied, designated ADH A, was in fact one half of the bifunctional enzyme externally referenced as AdhE, comprising an N-terminal aldehyde dehydrogenase (acetylating) and a C-terminal alcohol dehydrogenase. Overall it converts acetyl-coenzyme A to ethanol via ethanal as an intermediate; the bifunctional enzyme is the final step in the biosynthesis of ethanol in *G. thermoglucosidasius.* As the ADHE was under investigation for the purpose of increasing understanding about the native ethanol production system in *G. thermoglucosidasius,* previous work concentrated solely on ethanal as a substrate (Extance, 2012).

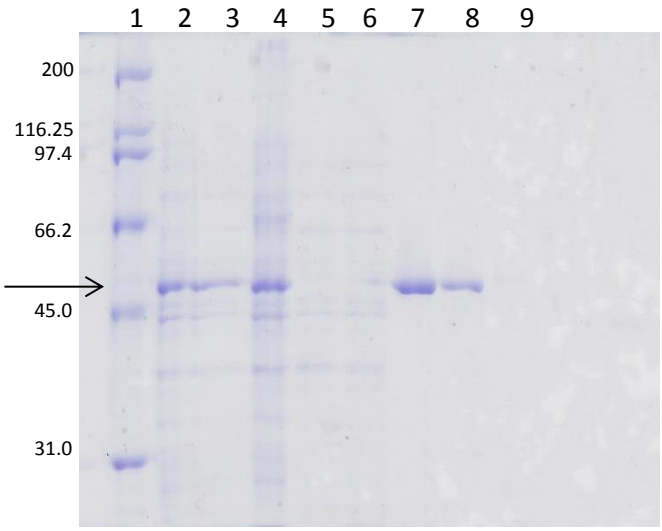


Figure 16 - SDS-PAGE gel of ADH A fractions post purification from IMAC chromatography column. Lane 1: markers, lane 2: total protein, lane 3: soluble fraction, lane 4: insoluble fraction, lane 5: 2nd flowthrough, lane 6: 0% His-Elute Buffer, lane 7: 10%, lane 8: 30%, lane 9: 100% His-Elute Buffer. The numbers on the left indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 48.6kDa.

2.1.1. SUBSTRATE SCOPE

Initial work was with disrupted cell extracts of *E. coli* BL21 (DE3) pET28a*,* in order to determine if activity was present. Whilst activity was confirmed with ethanal (data not shown), both cell extract and certain substrates absorb at 340nm, the wavelength under examination, causing difficulty. Additionally, even through judicious use of control experiments, interference by host alcohol dehydrogenases can never be entirely ruled out, and indeed this has been shown to be a major problem in previous work (Atsumi et al., 2010; Jarboe, 2011). Therefore, the decision was made to proceed with purifying the soluble fraction of the cell extracts, using Immobilised Metal Affinity Chromatography (IMAC), and using individual fractions thereof to

reduce the possibility of interference. This was facilitated by the N-terminal hexa-histidine tag provided by the vector pET28a.

A substrate screen was then carried out to determine the substrate scope of ADH A, with regards to its previously determined activity with ethanal. In the first instance substrates were screened on a volumetric rather than concentration basis.

Substrate	Relative Rate of Reaction (%)
Ethanal	100
Propanal	94
Butanal	78
Pentanal	15
Hexanal	19
Heptanal	29
Octanal	39
Nonanal	24
Decanal	10
Undecanal	5
Dodecanal	3
2-Pentanone	-
2-Hexanone	-

Table 1 – Relative rate of reaction obtained using ADH A with 10µl substrate in a 1ml assay. Additional conditions: 0.2mM NADH, 50mM citric acid, pH 6.0, 50 °C.

The lack of substrate solubility makes it difficult to compare rates between substrates such as ethanal, which is effectively miscible, with sparingly-soluble lengthier substrates. Specifically, by using 10µl substrate, the concentration ranged from approximately 138mM to 45mM. This variation in itself would be problematic, but pentanal and larger substrates have low solubility in water, of the order of 10mM and below. This means that it is not possible to accurately ascertain the real concentration of the substrate available to the enzyme, and the value of the experiment is limited.

It is worth noting though that if “excess” substrate is provided to an assay, invoking bilayer formation, then the substrate concentration will become limited to the solubility. Given that this is probable in this case, and that solubility decreases in a relatively linear manner with increasing molecular mass, then a concomitant reduction in specific activity would be expected if this were the sole factor. As this is not shown in the results, then there is clearly a confounding effect involved as the apparent rate of reaction does not reduce in a predictable fashion. For example, there is an increase in the rate of reaction when using heptanal as substrate instead of hexanal. If solubility was the limiting factor on the apparent rate, then having less substrate present in the aqueous phase would have resulted a lower rate of reaction when using heptanal. This could be explained by substrate inhibition, where the reduced solubility relieved an inhibitory effect caused by excess substrate.

All subsequent analysis was conducted on a substrate concentration basis, rather than by volume. In addition, where substrate solubility was known, experiments involving that particular substrate were kept below that critical value, treating experimental buffers as pure water for this purpose. Finally, smaller substrates were most frequently used in method development and analysis, both to ameliorate the experimental conditions, but also due to the apparent higher affinity of ADH A for the smallest substrates (i.e. ethanal, propanal, butanal).

The wide substrate scope with regards to aldehydes, but none with ketones, is not unexpected. Due to the wide conservation of this particular enzyme and its particular relevance to the production of ethanol, it has been widely studied, and this is a notable characteristic of this enzyme (Extance, 2012). That said, activity has only previously been noted for shorter aldehydes, up to octanal, and not beyond. This could be an anomaly peculiar to *G. thermoglucosidasius*, but more likely an artefact of examining the whole enzyme. The aldDH may be limiting, or it could be the scarcity of the larger substrates, e.g. decyl-coenzyme A.

2.1.2. KINETICS

Once the substrate scope was determined, the kinetic constants of the enzyme were considered. Due to the aforementioned issues, smaller substrates were preferred for this task, particularly as they would allow for a large range of substrate concentrations to be studied without the necessity of solvents.

To provide adequate analysis, both Hanes-Woolf and Michaelis-Menten curves were calculated, the former indicating the validity of the default assumption of adherence to Michaelis-Menten kinetics. Where necessary, both graphs are fitted assuming substrate inhibition using Equation 3 below instead of Equation 2.

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

Equation 2 – Standard Michaelis-Menten equation

$$v = \frac{V_{max}[S]}{K_M + [S] + \left(\frac{[S]^2}{K_i}\right)}$$

Equation 3 – Michaelis-Menten equation with substrate inhibition

Where:

v is the rate of reaction

V_{max} is the maximum rate of reaction

[S] is the substrate concentration

K_M is the Michaelis constant

K_i is the inhibition constant

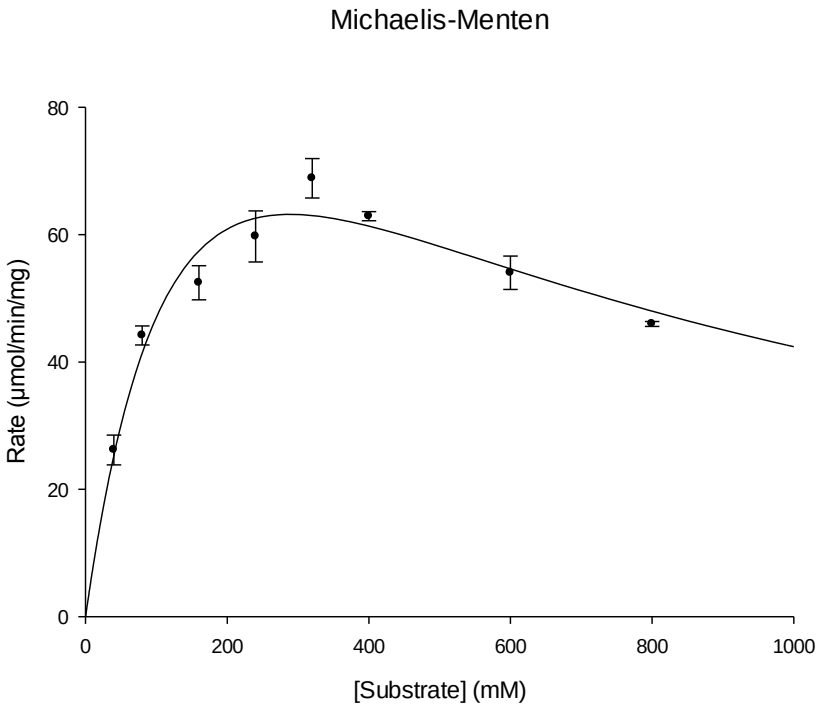


Figure 17 – Graph of rate of reaction vs [S] using ADH A and ethanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

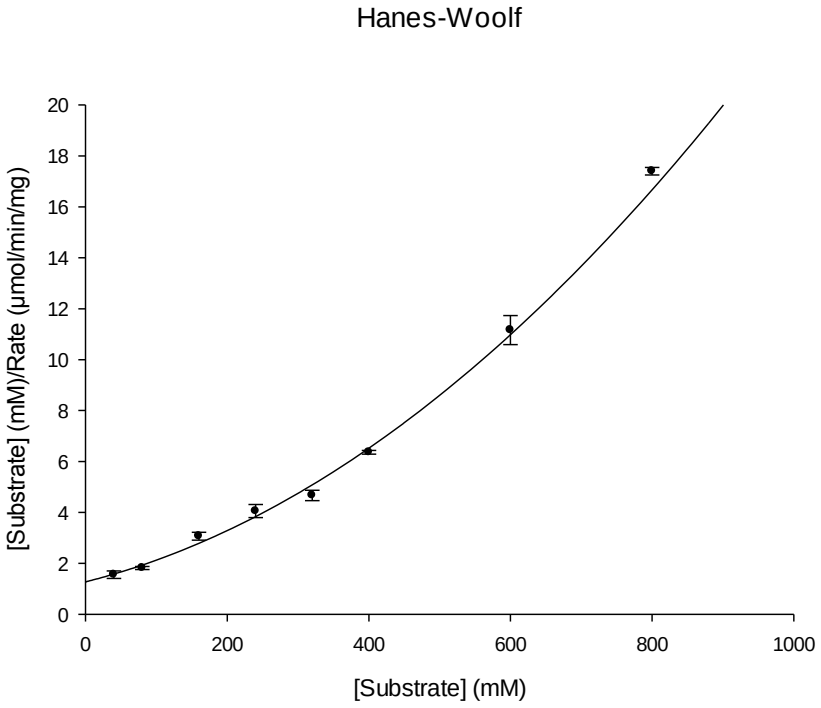


Figure 18 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and ethanal, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

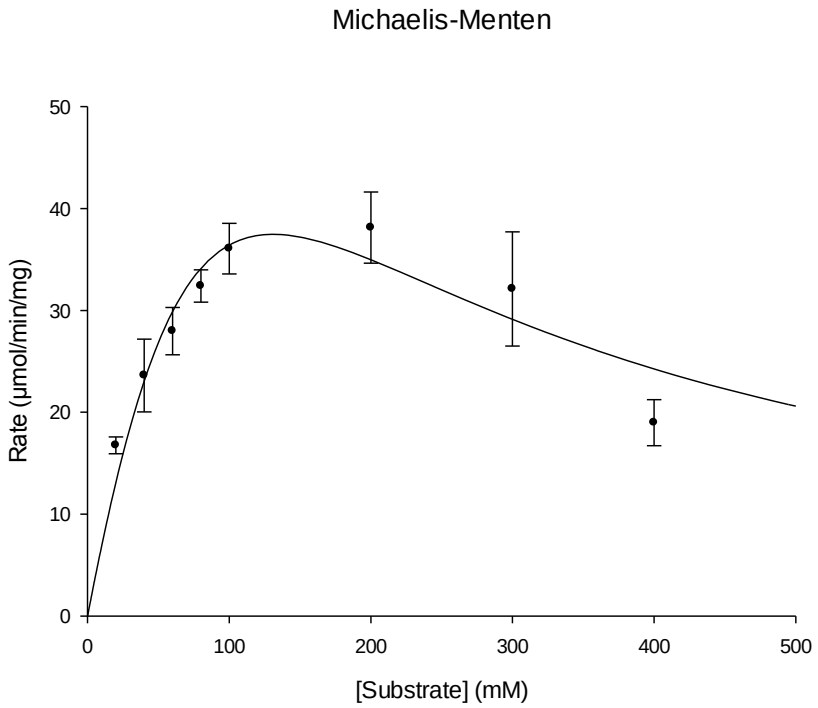


Figure 19 - Graph of rate of reaction vs [S] using ADH A and propanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

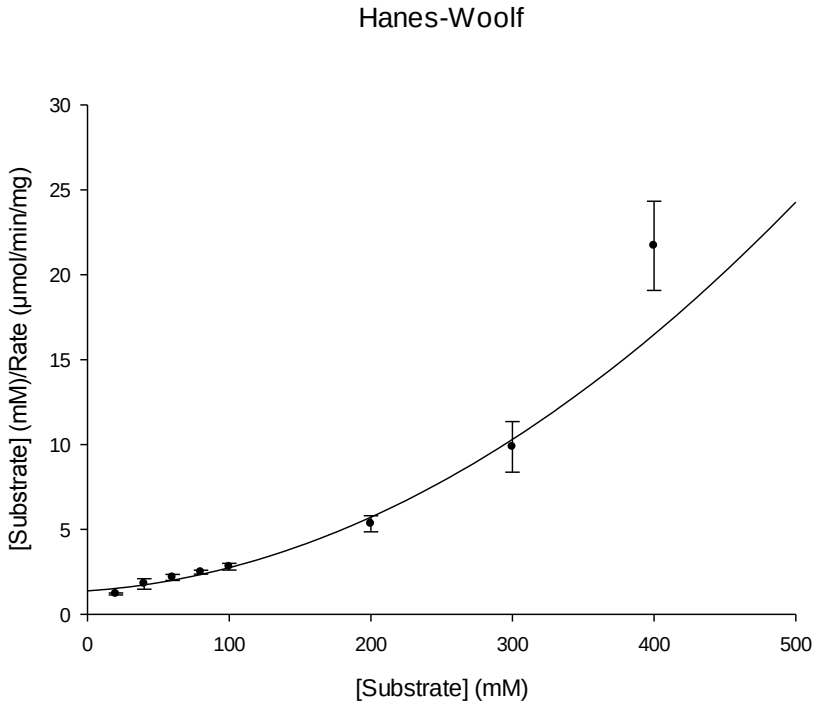


Figure 20 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and propanal, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

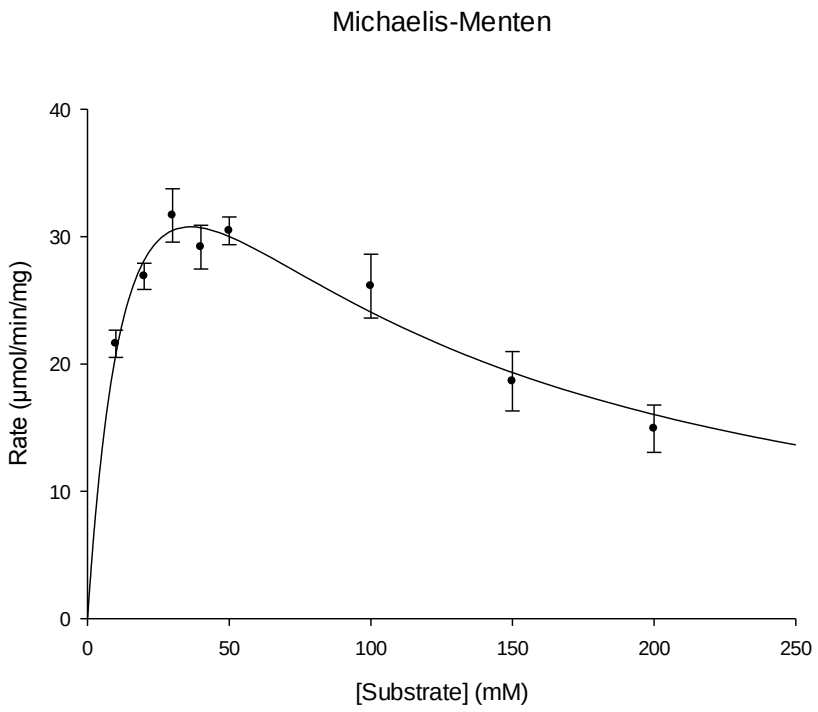


Figure 21 - Graph of rate of reaction vs [S] using ADH A and butanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

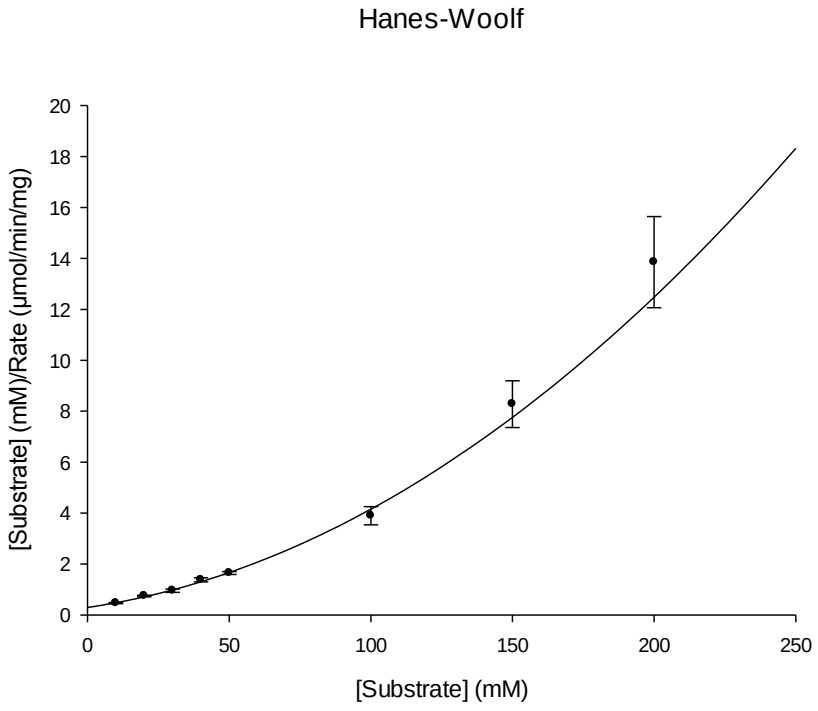


Figure 22 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and butanal, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

In all figures error bars represent the standard error, typically derived from triplicate experiments. The kinetic constants for ADH A are as follows:

Substrate	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
Ethanal	143 (±30)	182(±61)	456(±170)	7000	38
Propanal*	178 (±196)	246 (±326)	70 (±94)	8700	35
Butanal	62(±12)	18(±7.0)	73(±24)	3000	167

Table 2 - Kinetic constants calculated For ADH A. *Very high standard error; data should be considered illustrative only.

The Hanes-Woolf plots and the Michaelis-Menten curves indicate a lack of co-operative kinetics and that some level of substrate inhibition is present, as indicated by the decrease in enzyme velocity at higher substrate concentrations; the inhibition has contributed to a rather large standard error associated with the kinetic constants, it may have been helpful to have additional points at both higher and lower substrate concentrations, but particularly at the higher end.

Inhibition leads to a higher than expected V_{max}, which in turn leads to higher K_M, and therefore increased error as it is necessary to have data at substrate concentrations that exceed 3 fold higher than the K_M, even 10 fold higher in an ideal situation. Therefore, with a surprisingly high K_M, even with a very wide range of substrate concentrations, higher concentrations are liable to be needed to reduce the error. This level of error is compounded in the case of the K_i values recorded, simply because these values are typically higher than the equivalent K_M, and consequently require greater substrate concentrations to be fully elucidated.

Despite the errors noted it appears that ADH has a much improved affinity for butanal than ethanal, which was presumed to be its preferred substrate. The order of magnitude reduction in K_M is also intriguing because it would seem to indicate a preference for longer chain substrates, rather than a tolerance. The specificity constants calculated reinforce this belief, with an overall four-fold increase for butanal over both ethanal and propanal. This indicates that with a mixture of substrates at the same concentration, the rate will be four-fold higher with respect to butanal, even though a larger rate could in theory be obtained with either of the two other substrates (Eisenthal et al., 2007).

2.1.3. STABILITY ANALYSIS

In addition to substrate scope, an investigation into the stability of ADH A was conducted, primarily considering solvent stability. DMSO, acetonitrile and hexane were the principal solvents tested, as they are frequently used industrially as cosolvents for biocatalytic steps. Taken together they represent a range of solvents, due to their miscibility or otherwise with aqueous solutions. In particular hexane is predominantly forming bilayers which may prove useful to partitioning substrate and product from the enzyme, whilst DMSO is most likely to improve solubility of both in the aqueous phase.

Using octanal as substrate and DMSO as solvent, 1vol% solvent caused a reduction in observed activity of 6%. With acetonitrile, 14% reduction was noted, and with hexane 10%. Increasing the proportion of hexane proved difficult as it attacked both the cuvette and the parafilm used for covering the cuvette whilst inverting. Further, due to immiscibility, the absorbance when using hexane was exceptionally high, resulting in inaccurate or indeterminate readings.

With regards to acetonitrile though, increasing to 10vol% and 20vol% resulted in a 75% and 93% reduction in observed activity respectively, indicating substantial inactivation with increasing solvent. ADH A was more resilient when presented with DMSO, however, with a 92% reduction in activity only occurring with the addition of 50vol% DMSO. Whilst it is unlikely that an industrial environment would require a 50% DMSO cosolvent, due to the expense and the difficulty in separating the solvent from water, it is still useful to know where the boundaries lie.

If such a high volume of DMSO is required for solvating the substrate and/or product, then it would probably be better to consider a biphasic reaction, with either the substrate dissolved in hexane or similar, or a solvent free system where the substrate itself forms the second phase (Straathof et al., 2002; Doukyu and Ogino, 2010). Typically, if a high concentration of solvent is to be used, then it is as an alcohol as a sacrificial substrate, or to work in anhydrous conditions. Other options include non-standard solvents and ionic liquids to improve solubility (Schroer et al., 2007; Musa et al., 2008; Gu and Jérôme, 2010). More exotic routes such as gas/solid reactors also exist (Kulishova et al., 2010). High concentrations of solvent may also affect the structure of enzymes in unexpected ways, not just in terms of stability. The substrate scope may be affected, as well as enantiospecificity and the enantioselectivity. This effect is particularly noted at near anhydrous conditions, when the enzyme is either lyophilised or the reaction is carried out in pure solvent.

Further to considering specific activities in the presence of solvents, a kinetic analysis was conducted in the presence of 10vol% DMSO to determine if the kinetic constants were affected.

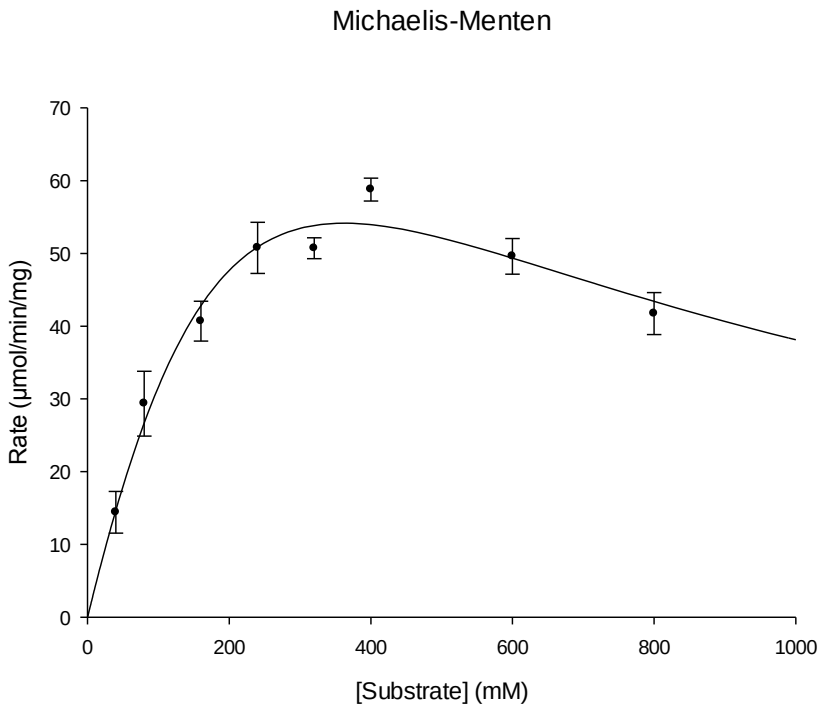


Figure 23 - Graph of rate of reaction vs [S] using ADH A and ethanal in the presence of DMSO. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume, 10vol% DMSO.

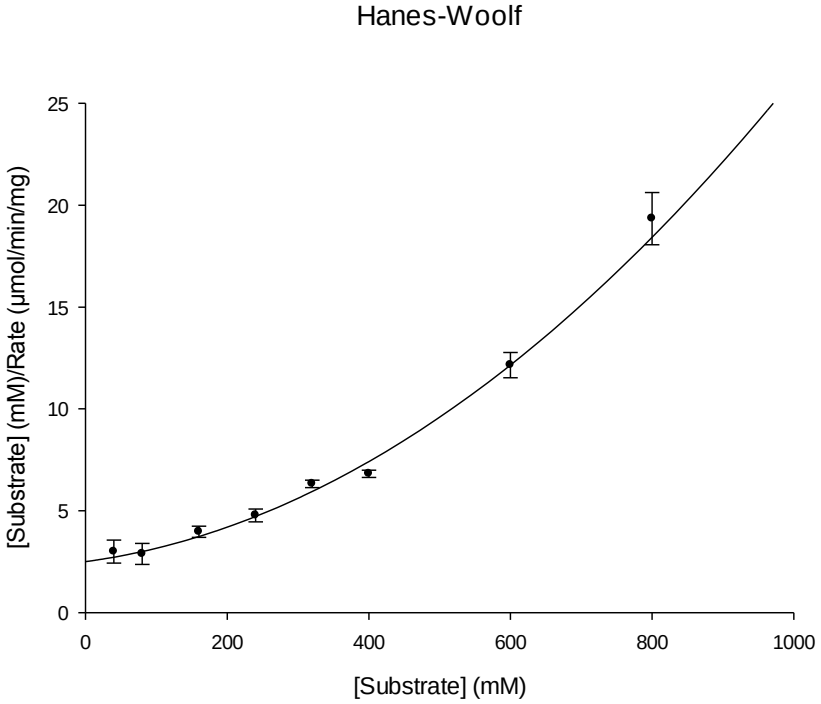


Figure 24 - Hanes-Woolf plot ([S]/v vs [S]) for ADH A and ethanal in the presence of DMSO, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume, 10vol% DMSO.

It is not possible to say definitively whether the addition of DMSO has affected the kinetic constants due to the standard error present, again due to the requirement for further data at substrate concentrations of 1M and above. By visual inspection, however, there appears little to choose between Figure 17 and Figure 23; this cannot be shown statistically. The constants obtained using DMSO are: V_{max} 216 U, K_M 40 mM and K_i 243 mM. It is also worth noting that it can be seen from Figure 24 that Michaelis-Menten kinetics is maintained even in the presence of the solvent.

2.1.4. SEQUENCE ALIGNMENT

To consider the experimental evidence obtained with reference to previous knowledge, the amino acid sequence of ADH A was aligned with similar sequences and key amino acids highlighted. Given the availability of a crystal structure for the ADH section of the full enzyme, this was something also to be investigated; however, comparing similar sequences became rather complex when attempting to align ADH sequences and whole enzyme sequences, so two alignments were carried out. In the first multiple sequence alignment, the nearest matches for the entire enzyme were aligned with the whole sequence (ADHE). All these similar sequences had 65% or greater sequence identity with ADH A. This contrasts with the alignment of the ADH section only, which was aligned with the most similar sequences held in the Protein Data Bank (PDB), with these sequences typically having 30% or less sequence identity.

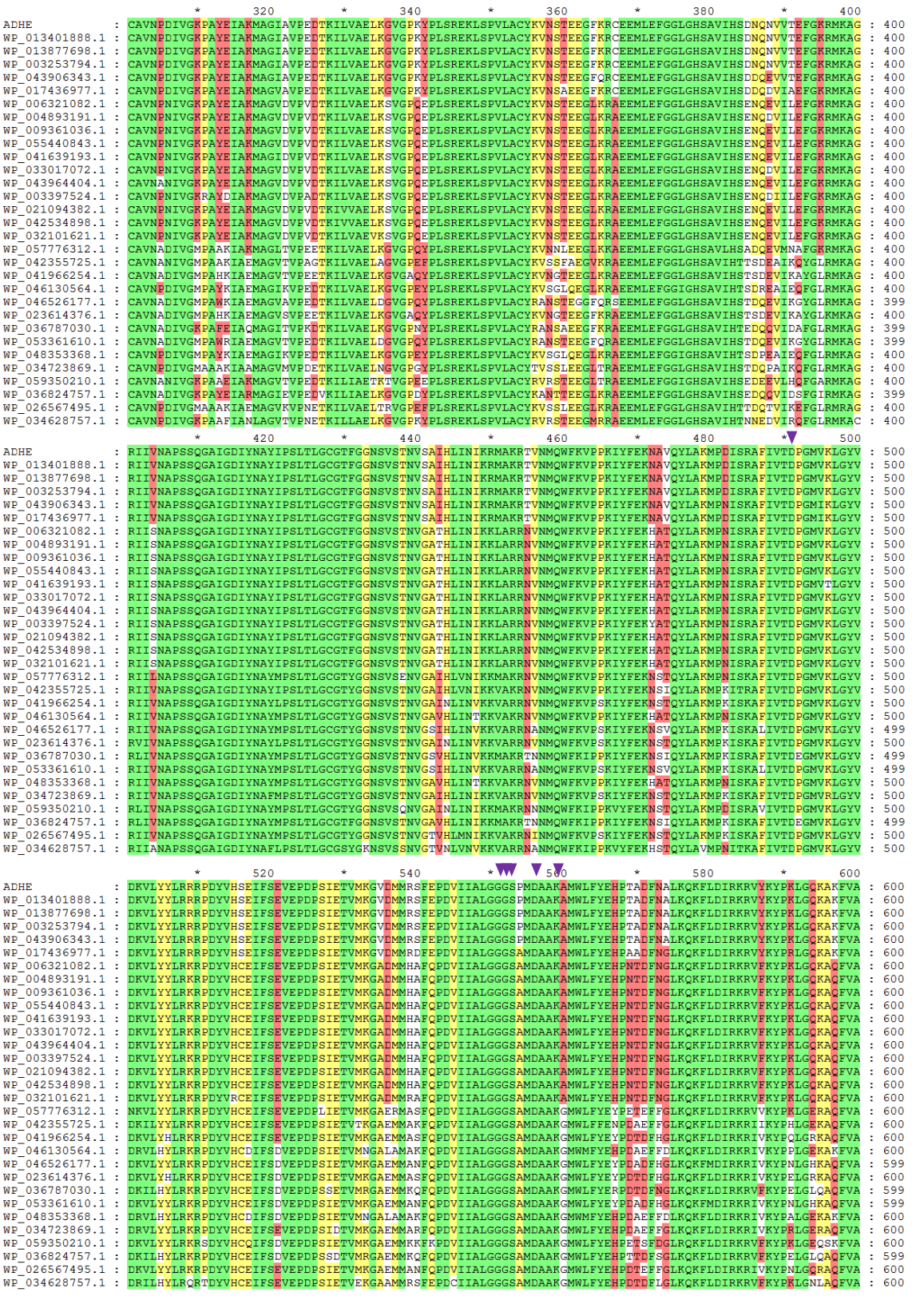


Figure 25 - Multiple Sequence Alignment (MSA) of *G. thermoglucosidasius* ADHE enzyme with sequences identified using NCBI Seqr Sequence Search. The blue triangle indicates the putative catalytic cysteine residue, purple triangles indicate other putative active site residues, and grey triangles indicate the metal-binding site, all obtained from the NCBI Conserved Domain Database via the Conserved Domain Search. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to NCBI protein sequences from the Reference Sequence (RefSeq) database. Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997). All sequences are annotated as bifunctional acetaldehyde-CoA/alcohol dehydrogenases and are from the genera *Geobacillus, Anoxybacillus, Bacillus* and *Pontibacillus*.

What is striking about this alignment is the very high level of conservation. Virtually every single residue is conserved in at least 15 of the 30 closest sequences; only 25 of the 867 residues are not. Given that these enzymes are from the same family and ADHs from the same genus are often very similar (>90% sequence identity) this is probably not unusual (Radianingtyas and Wright, 2003).

The least conserved area of the bifunctional enzyme is in the initial 40 amino acids, although there are areas throughout the enzyme which are less conserved, e.g. positions 180-200 and 820-840. With such high levels of conservation there are few positions where there is any difference in any of the highlighted residues – D659Y for WP_032101621.1, from *Anoxybacillus flavithermus*, annotated as a metal binding site; Y626I for WP_059350210.1, from *Bacillus coahuilensis*, annotated as part of the active site; H740G for WP_046526177.1 and WP_046526177.1 from two *Bacillus* spp and also annotated as part of the active site. In all three cases the amino acid change is liable to have an effect as only in the case of Y626I are both amino acids similar, in this case both classed as hydrophobic. That said, tyrosine is substantially more bulky than isoleucine, which may have an effect.

In all these cases there could be a substantial effect on either the substrate profile of the enzyme or the kinetics. With the modified residue believed to interact with the metal ion, this could also impact upon the stability (e.g. thermostability) in addition to catalytic effects. As no further information exists on the NCBI database with respect to these proteins, this remains speculative.

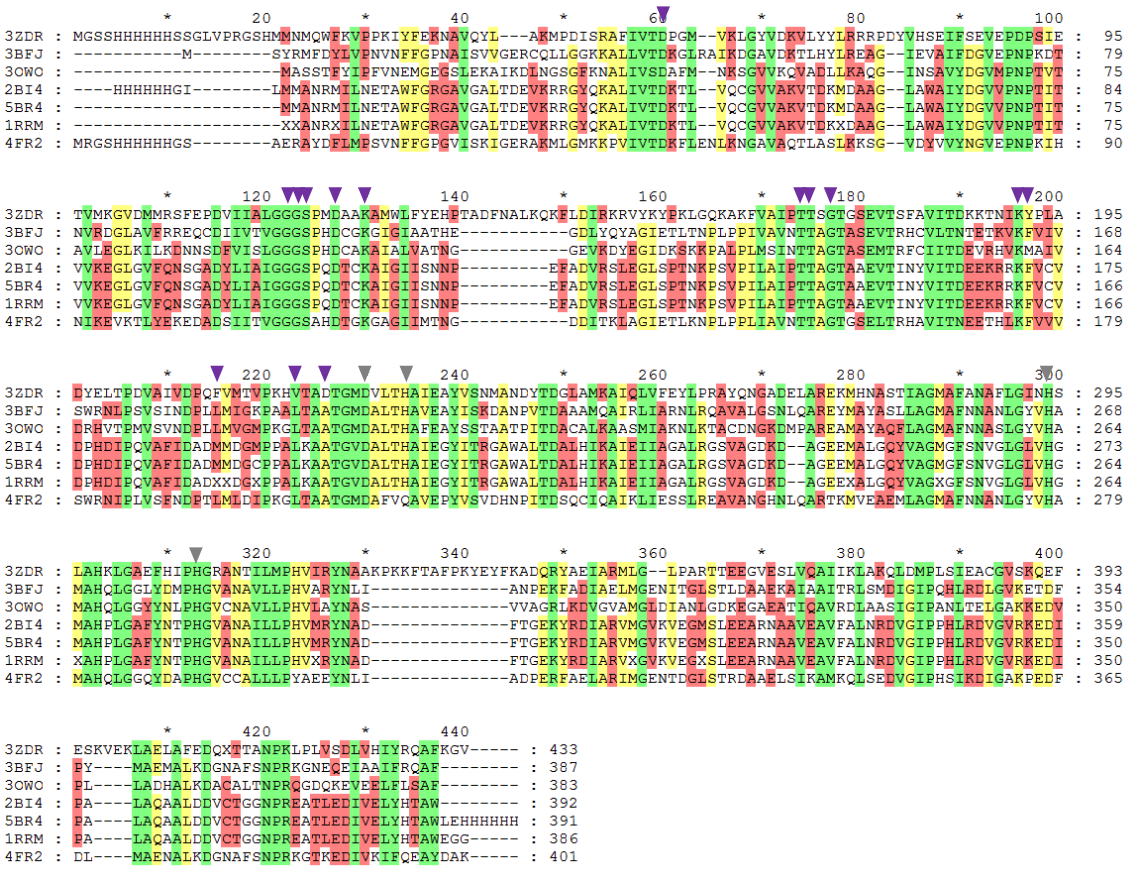


Figure 26 - Multiple Sequence Alignment (MSA) of ADH A (the C-terminal ADH domain of ADHE), using the sequence obtained directly from crystal structure 3ZDR. Purple triangles indicate putative substrate binding site, and grey triangles indicate metal binding site, obtained from the NCBI Conserved Domain Database via the Conserved Domain Search. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to sequences from PDB with associated crystal structures. Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997). Sequences are annotated as follows: 3BFJ – 1,3-propanediol oxidoreductase from *Klebsiella pneumoniae*, 3OWO – Iron dependent ADH from *Zymomonas mobilis*, 2BI4 – Lactaldehyde: 1,2-propanediol oxidoreductase from *E. coli* (FucO), 5BR4 – Lactaldehyde reductase from *E. coli*, 1RRM – Lactaldehyde reductase from *E. coli*, 4FR2 – ADH from *Oenococcus oeni*

Looking more broadly, the second multiple sequence alignment (MSA) considers the closest amino acid sequences held on the Protein Data Bank (PDB). In contrast to the >65% sequence identity in the previous MSA, these sequences have 32-35% sequence identity with 3ZDR, the sequence reported as part of the crystal structure obtained of the ADH portion, ADH A.

As can be seen in Figure 26, despite a drop in conserved residues, few changes are made to residues noted for their role in either the substrate binding site or the metal binding site. Y197 is the first, being generally conserved as phenylalanine, but once (3OWO) being replaced by methionine instead, a substantially different amino acid. The only other different key residue is F215, which is replaced variously with methionine and lysine. Both these positions are believed to be related to the substrate binding site and therefore may have a substantial effect. In addition, ADH A (3ZDR) has two areas (positions 140-150 and 330-345) that are unique. These are believed to be due to the bifunctional nature of the full enzyme, and therefore are understood to be related to the stabilisation of the full enzyme. As the remainder of the sequences are ADHs only, they do not have these sequences (Extance et al., 2013).

3BFJ may be a 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae*, but its structure is noted to be similar to other ADHs and in particular a member of the iron-containing ADHs (Ma et al., 2010). It also forms a decameric quaternary structure, which although is not in the form of a spirosome, it is nonetheless a substantial structure (Marçal et al., 2009). Despite this difference in function, additional investigations with 3BFJ may be of note. It has been discovered that mutating a single residue (D60G or D60A by numbering above, D41 original sequence) converts the enzyme from only utilising NADH to one able to utilise both cofactors. In the case of D60G, NADPH was preferred to NADH by approximately 1.5 fold, by specificity constant, with a 15-fold preference for NAD^+ over $NADP^+$ (Ma et al., 2010).

3OWO is an NAD^+ dependent, ethanol producing, oxygen sensitive ADH from *Zymomonas mobilis*. It is a dimeric enzyme of the iron-containing ADH family implicated in the cytosolic respiratory system. Despite the dimeric form, its structure is noted as similar to other ADHs (Moon et al., 2011).

2BI4 is from *E. coli* one of the first members of the iron-containing ADH family to be identified (Conway and Ingram, 1989) and is involved in the conversion of lactaldehyde to 1,2-propanediol. It is specific for *S*-enantiomers of the diol and aldehyde and is NAD(H) dependent. It also has activity with a range of small aldehydes and alcohols, although K_M values range from high to very high in all cases indicated, and k_{cat} figures reported are very low (Blikstad and Widersten, 2010). The limitation on substrate size is believed to be due to the tertiary structure, which has a "narrow waist", limiting the size of the substrate admitted to the active site. Further, the same residue as above (D60, native D39) has been determined to be key for cofactor specificity (Montella et al., 2005).

Lastly, 4FR2 is an NADH-dependent dimeric ADH of the iron-containing ADH family from *Oenococcus oeni*. Whilst the primary substrate was believed to be ethanol, other substrates such as 1,2-propanediol and 1,3-propanediol were also accepted, although the latter required the addition of $NiCl_2$ and specific activities noted were below 2 U mg^{-1} (Elleuche et al., 2013).

Overall, the most similar amino acid sequences for the full bifunctional enzyme appear to have some discrepancies in the key residues believed to be involved with the active site and metal binding site, which may have implications for their activity and stability. With regards to the more distant sequences, but most similar obtainable from the PDB, all these sequences are iron-containing ADHs, but may be more specifically annotated, e.g. as 1,3-propanediol dehydrogenase. There is some debate as to whether all these enzymes have iron or zinc as their metal ion. ADH A has zinc, but some of the other enzymes differ in this regard, which is in keeping with the belief that this family of enzymes are more accurately described as metal-dependent rather than iron-dependent. From work done on other enzymes, it appears that one key residue is D61, which is crucial to the cofactor dependence of the enzyme. In addition, structural work has indicated that the other residues noted in the MSA are likely to be important for activity across all enzymes in the MSA, but given the occasional differences, this may not always hold true. Further, given the substantially different substrate profiles and activities it may not be possible to directly compare these enzymes in terms of their function, but solely in terms of their structure.

2.1.5. SUMMARY

From the data obtained it appears that ADH A not only has a larger substrate scope than previously shown in similar enzymes, but may prefer larger substrates than previously believed. This has important implications for continued study of this ADH and the overall bifunctional enzyme, ADHE, and for the production of higher alcohols via biosynthesis.

If it can be determined whether the AldDH can perform as well as the ADH, then this could potentially be a route to the biosynthesis of higher alcohols if suitable precursors can be generated *in vivo*. These precursors would probably have to derive from fatty acid biosynthesis and therefore would be a complex pathway to manipulate.

Another potential issue is that it is believed that large modifiers such as coenzyme A are designed to reduce K_M for smaller substrates and to increase specificity downstream i.e. in those enzymes where coenzyme A-modified substrates are accepted, the use of the modifiers enhances the chances of those enzymes recognising their true substrates. The drawback of this system is that the coenzyme A-accepting enzymes *de facto* recognise the coenzyme A rather than the small substrate and therefore become promiscuous with respect to the small substrate. That said, this system is considered easier to develop and maintain natively than having several enzymes in a pathway exclusively recognise a small substrate without modification (Bar-Even et al., 2011). If this were true, however, it could be favourable in that the AldDH would be more inclined to recognise other coenzyme A-modified substrates, even if its affinity would be lower.

Speculatively, it would also imply that larger substrates could be imposed upon such enzymes, such as the AldDH, if appropriate modifications to the active site were made. Presumably the major reason for reduced affinity of larger substrates is solely physical size of the substrate itself, even with coenzyme A having a molecular mass of 767, so the possibility exists of engineering a more amenable active site. Without a crystal structure it is difficult to make a judgement; however, coenzyme A is a relatively linear molecule, thus it seems more likely the length of the substrate may become an issue, not fitting neatly in the active site, rather than a question of the substrate being able to access the active site at all.

Finally, it should be emphasised that ADH A is one half of a bifunctional enzyme, which means that its performance may not be similar in its native state. In particular, the notably high K_M values suggested by the recombinant ADH may not be indicative of the behaviour of the native enzyme *in vivo*, and care should be taken when attempting to interpret these results in this manner. What is more telling is the relative change between substrates; that is the key result obtained in this work.

2.2. ADH B

This enzyme was under investigation to produce 2-butanol from 2-butanone, as a final step of an *in vivo* process shown in Figure 27. As such, it was already known to be a ketone utilising ADH and NADH dependent.

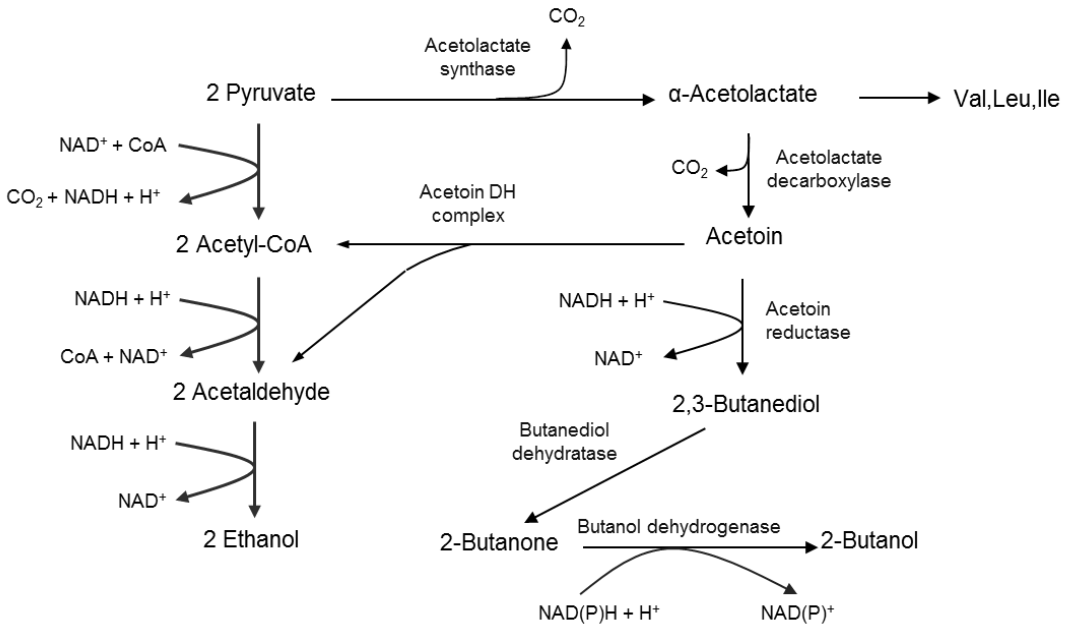


Figure 27 - Butanol pathway under investigation. Courtesy of Dr Leann Bacon and Dr Steve Bowden (University of Bath, UK), partially reproduced from Cripps et al. (2009).

On this basis the enzyme was investigated for its potential substrate scope to identify additional substrates of interest. Further, this enzyme was used substantially in methodology development, particularly in determining appropriate substrate concentrations to ameliorate the quality of kinetic data obtained. Another principal alteration to the method at this time was a switch from polystyrene (PS) cuvettes to poly (methyl (methacrylate)) (PMMA) cuvettes, for improved solvent and substrate resistance.

No work was conducted using disrupted whole cells with this enzyme or any further enzymes examined, only purified fractions were tested. Purification of the enzyme was conducted using IMAC, and individual fractions produced were assayed with both ethanal and 2-butanone in the presence of NADH. The N-terminal hexa-histidine tag was not removed prior to assay and raw fractions were stored as they were eluted from the column; no further purification or buffer exchange was done.

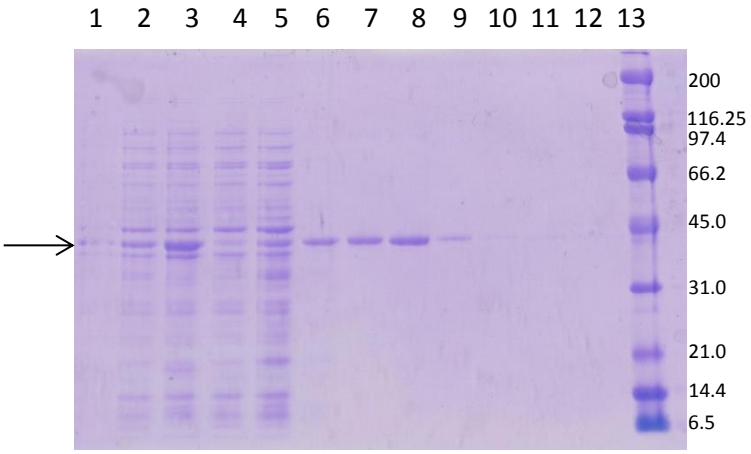


Figure 28 - SDS-PAGE gel of ADH B. Lane 1: total protein, lane 2: insoluble fraction, lane 3: soluble fraction, lane 4: 2nd flowthrough, lane 5: 1% His-Elute buffer, lane 6: 2.5%, lane 7: 5%, lane 8: 10%, lane 9: 20%, lane 10: 30%, lane 11: 50%, lane 12: 100% His-Elute buffer, lane 13: Protein markers; values provided on right indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 39.6kDa.

2.2.1. SUBSTRATE SCOPE

Additional substrates were sourced from literature as examples of those being investigated in the past and present due to their industrial relevance. In this manner simple aldehydes and ketones as well as complex substrates were utilised in the initial substrate screen.

Moreover alcohols were screened, using similar conditions as with the aldehydes and ketones, specifically using 50mM citric acid buffer at pH 6.0. There was one alteration, using 2.5mM NAD^+ rather than 0.2mM NADH. This was done to reduce the chance that cofactor limitation was interfering with the assays, a consideration that cannot be made when using NADH because of the high absorbance; typically 0.2mM or similar concentrations are used when assaying in the reductive direction.

Substrate	Specific Activity (10mM) [U mg^{-1}]	Specific Activity (20mM) [U mg^{-1}]
Ethanal	56.3	n.d.
2-Butanone	75.9	77.1
Acetoin*	32.9	50.1
2-Pentanone	24.7	52.5
2-Hexanone	1.6	1.5
3-Hexanone	0.4	2.2
Ethyl 4-chloroacetoacetate	90.7	80.6
6-Methyl-5-hepten-2-one	-	-
4-Phenyl-2-butanone	-	-
2,2,2-Trifluoroacetophenone	-	-
Ethyl-2-oxo-4-phenylbutyrate	0.9	-
Ethyl 4,4,4-trifluoroacetoacetate	0.1	0.2
Methyl benzoylformate†	-	-
2-Decanone	-	2.1
Propanal	79.0	81.3
Butanal	94.9	110.2
Pentanal	29.8	35.3
Hexanal	0.8	0.9

Table 3 - Initial substrate screen with ADH B. Substrates assayed at 10mM/20mM concentration unless otherwise noted. *Assayed at 1/2mM. n.d. indicates not determined,"-" indicates no activity noted in assay. † Substrate absorbs at 340nm, negating assay viability. Conditions of assay: 50mM citric acid, pH 6.0, 0.2mM NADH, 1ml assay volume.

Substrate	Specific Activity (10mM) [U mg^{-1}]	Specific Activity (20mM) [U mg^{-1}]
Ethanol	0.9	1.7
Propanol	4.5	6.9
Butanol	2.6	5.2
Hexanol	-	-
2-Propanol	53.5	53.2
2-Butanol	65.7	61.3
2-Pentanol	54.9	33.4
3-Pentanol	16.0	31.2
1-Phenylethanol	-	n.d.
1,2-Propanediol	8.3	18.6
Glycerol	-	n.d.

Table 4 – Alcohol screen using ADH B. n.d. indicates not determined,"-" indicates no activity noted in assay. Conditions of assay: 50mM citric acid, pH 6.0, 2.5mM NAD^+, 1ml assay volume.

On the basis of these screens it was possible to clearly demarcate the substrate scope for this enzyme, which was entirely based on size. With the sole exception of glycerol, all tested substrates of five carbons or less in length were accepted, with any larger having residual activity at best. No activity was noted with any aromatic substrate, which isn't overly surprising given the indications of a near absolute cut off in substrate scope after five carbons.

2.2.2. INITIAL KINETICS

After establishing the substrate scope of ADH B, kinetics of various substrates were considered. With this enzyme, the decision was taken to focus primarily on the lower substrate concentrations, up to 5mM, to ensure that all substrates tested were significantly below the solubility limit. Additional reasoning is that the average enzyme has a K_M in this region (Bar-Even et al., 2011), with the median value of approximately 100µM and 60% of all K_M values recorded fall below 1mM. Thus a variety of substrates were chosen, and for the first time the kinetics of the co-factor were examined as well.

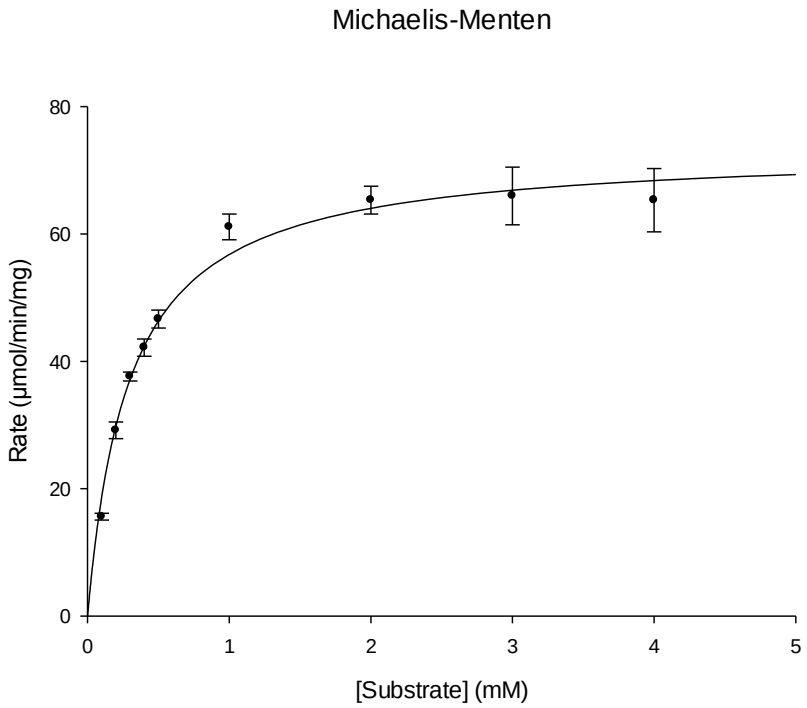


Figure 29 - Graph of rate of reaction vs [S] using ADH B and 2-butanone. Data fitted to a standard Michaelis Menten equation using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

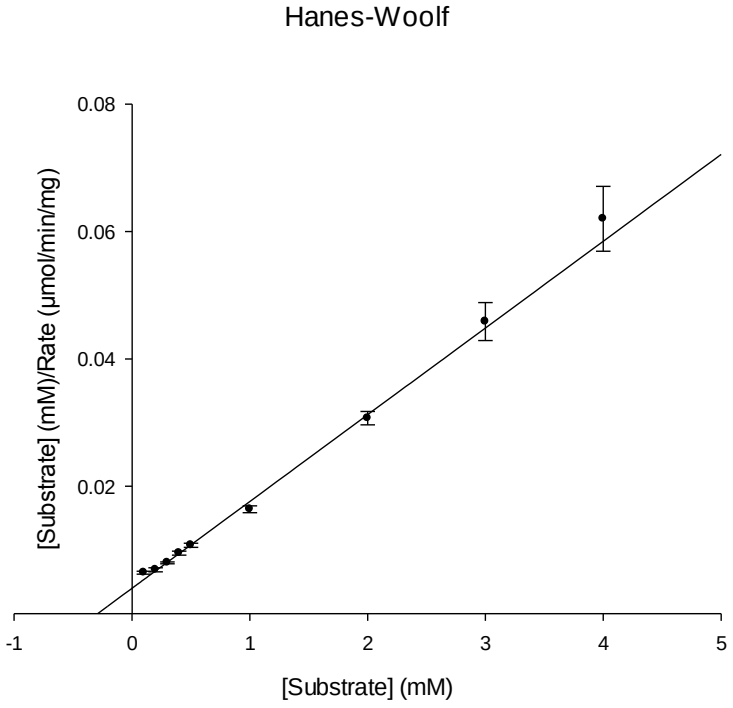


Figure 30 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

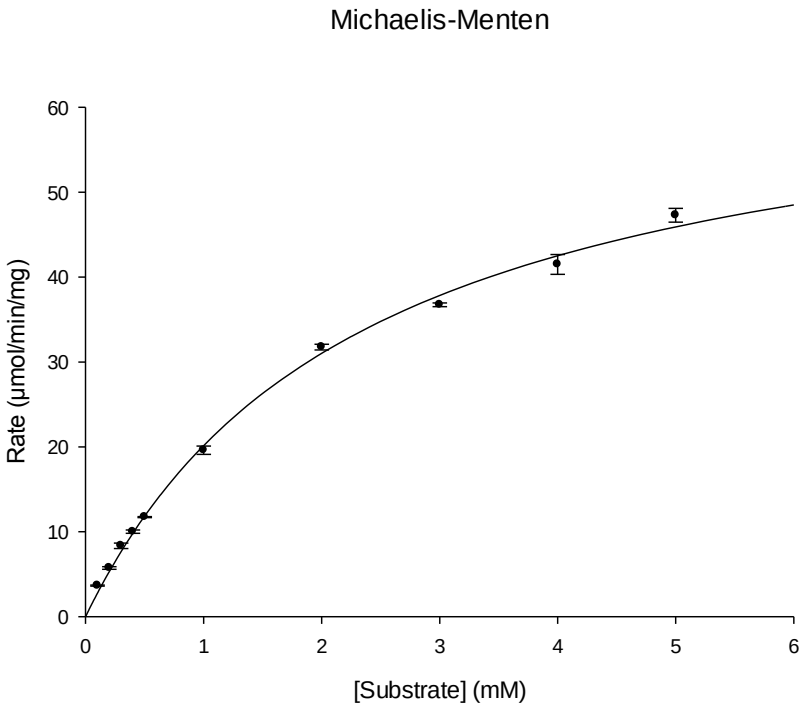


Figure 31 - Graph of rate of reaction vs [S] using ADH B and acetoin. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

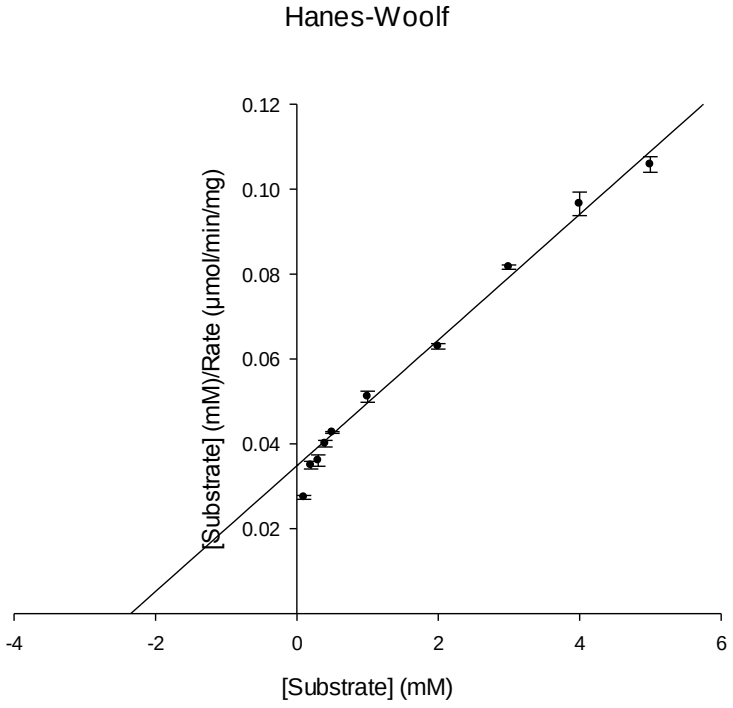


Figure 32 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and acetoin. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

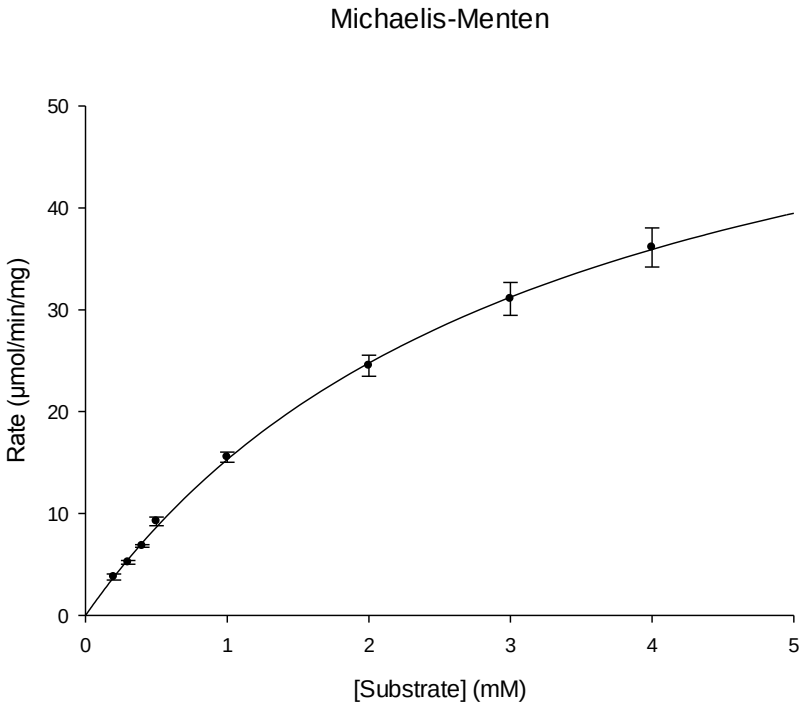


Figure 33 - Graph of rate of reaction vs [S] using ADH B and ethyl 4-chloroacetoacetate. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

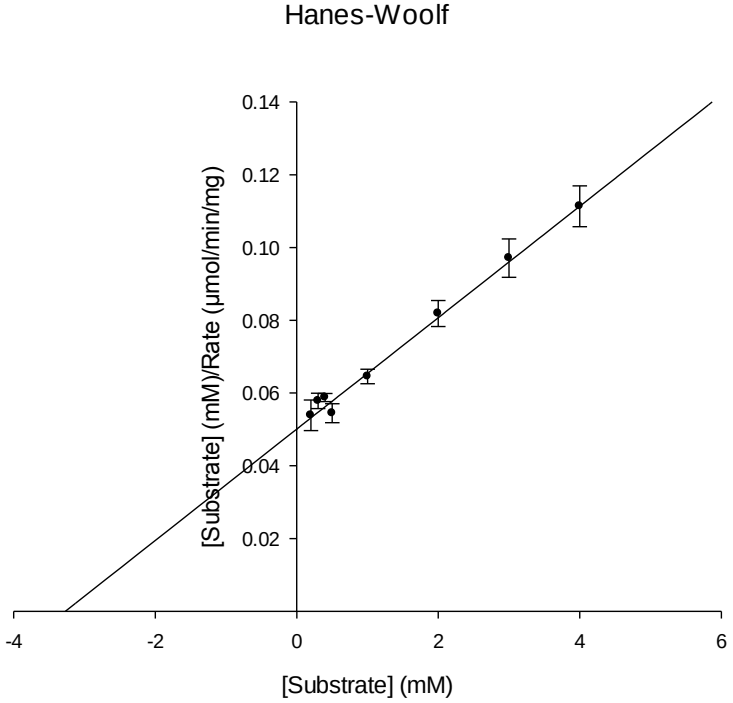


Figure 34 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and ethyl 4-chloroacetoacetate. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

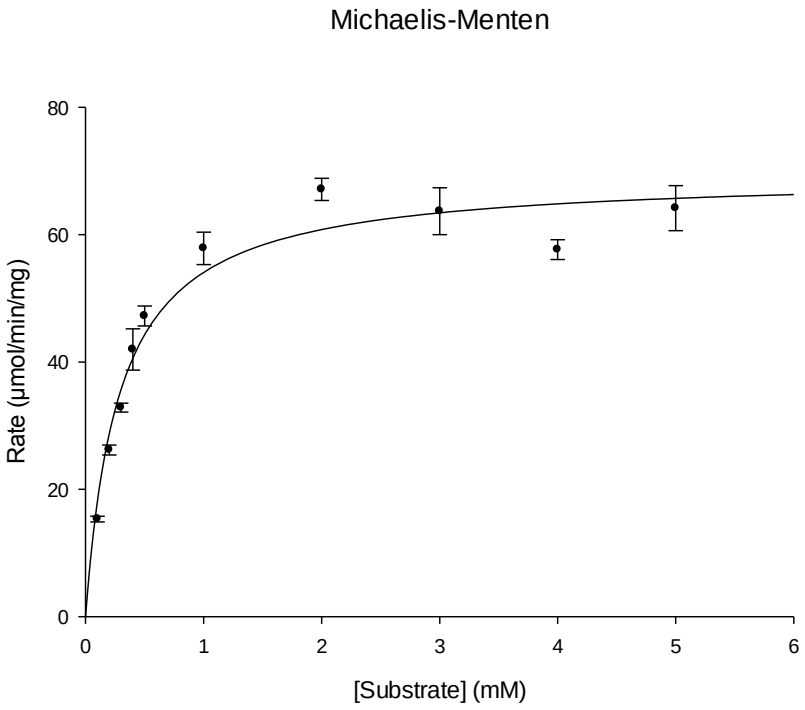


Figure 35 - Graph of rate of reaction vs [S] using ADH B and propanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

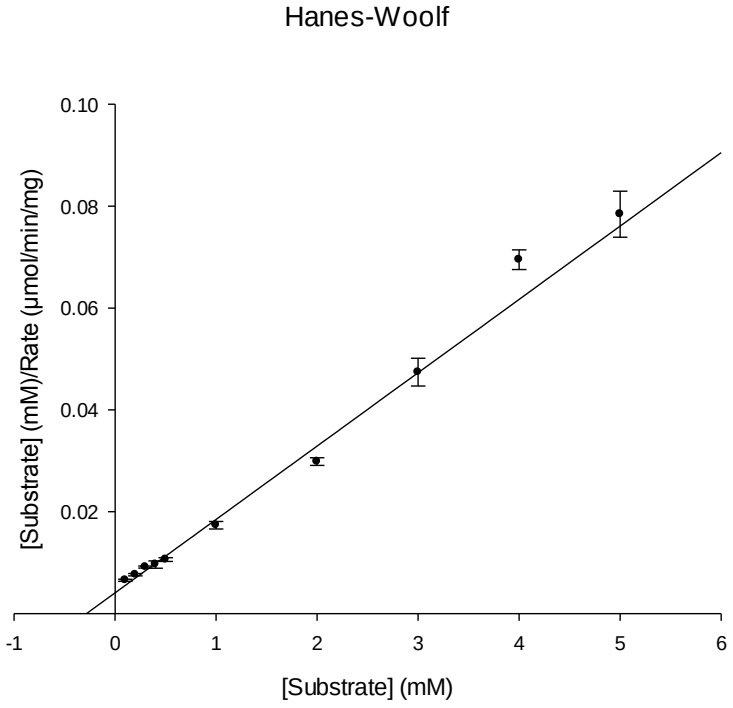


Figure 36 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and propanal. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

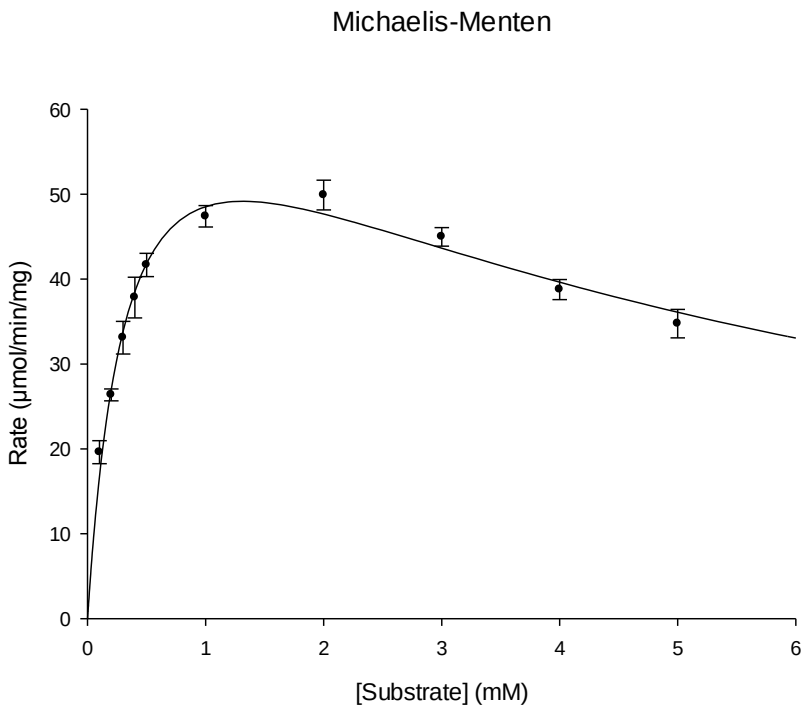


Figure 37 - Graph of rate of reaction vs [S] using ADH B and 2-propanone. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Figure 38 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-propanone, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

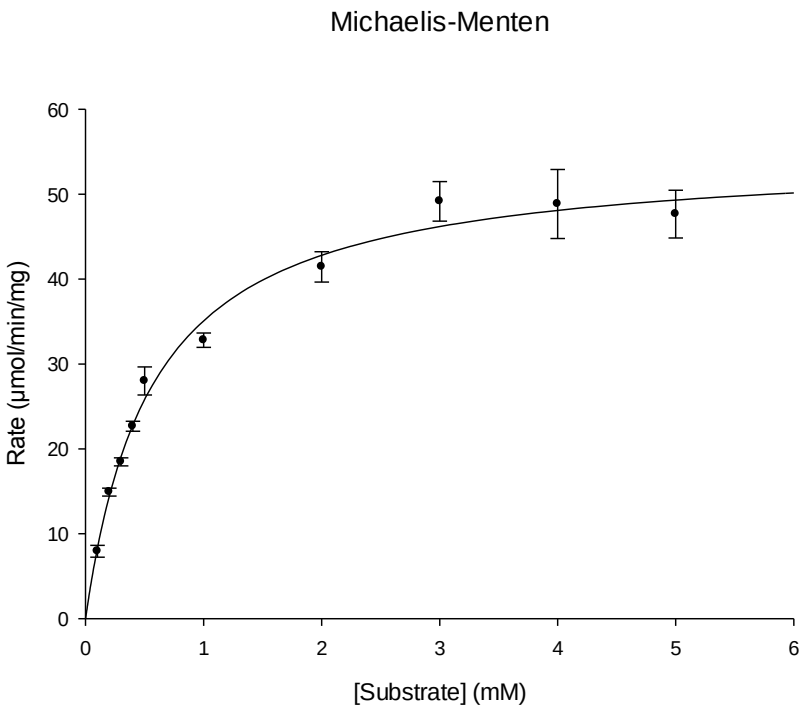


Figure 39 - Graph of rate of reaction vs [S] using ADH B and 2-butanol. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NAD^+, 50 °C, 1ml assay volume.

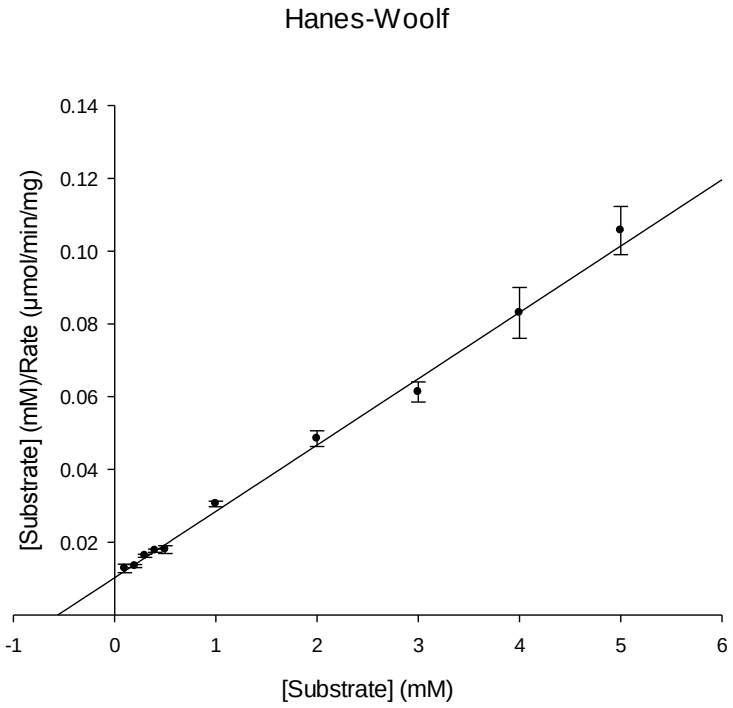


Figure 40 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanol. Conditions: 50mM citric acid pH 6.0, 0.2mM NAD^+, 50 °C, 1ml assay volume.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})	Michaelis-Menten Kinetics
2-Butanone	73.4±1.99	0.29±0.03	-	2900	10000	Yes
Acetoin	67.5±1.72	2.35±0.13	-	2700	1100	Probable
COBE*	65.3±4.62	3.27±0.41	-	2600	800	Yes
Propanal	69.4±1.95	0.28±0.03	-	2800	9800	Yes
2-Propanone	75.5±5.66	0.35±0.05	4.89±0.90	3000	8600	Yes, Inhibition
2-Butanol	54.9±1.61	0.56±0.06	-	2200	3900	Yes

Table 5 - Summary of kinetics obtained with ADH B. *COBE is ethyl 4-chloroacetoacetate. "-" indicates not noted.

There are several fascinating points about these results. Firstly, it is clear that 5mM is an insufficiently large substrate concentration to use to determine kinetic constants in this particular case. This can be seen in the specific activities reached for propanal (approximately 60-70 U mg^{-1}) and the estimated V_{max} of 70 U, as these are substantially below the 80 U previously noted with 10mM and 20mM substrate concentration. Whilst variation is to be expected, there is a large enough discrepancy to warrant slightly higher substrate concentrations wherever possible.

The similar V_{max} results contrast with a comparatively varied set of K_M constants obtained, with roughly an order of magnitude variation. On the basis of results thus far, it is not possible to state with any certainty whether 2-butanone, 2-propanone or propanal are the most preferred substrate, although substrate inhibition was noted with 2-propanone. Even the specificity constants are less than forthcoming, with 2-propanone lagging slightly behind both 2-butanone and propanal.

2.2.3. METHODOLOGY DEVELOPMENT

When conducting the substrate screen previously, raw substrate was used, typically ~1µl in a 1ml assay. This was done to allow for a rapid screen. In the case of kinetics, a stock solution was used to decrease the error between any given points by using the same stock solution for varied concentrations. This necessitates using a higher concentration stock solution than the concentration intended at assay conditions, which is usually where the issues of solubility come to the fore.

Solubility was not an issue at concentrations of 10mM and below, particularly with the substrates chosen. However, at this stage, the decision was taken to expand the range of substrate concentration from 50 fold to 800 fold. This was done by changing the stock solutions used to 100mM instead of 10mM and was carried out to provide additional data at higher substrate concentrations; despite the earlier belief that 5mM would be sufficient, this appeared not to be the case.

Additionally, during this work a steady reduction in recovered protein was noticed, ultimately resulting in virtually none being obtained. This correlated with several replacement probes for the sonicator used to disrupt cells and anecdotal evidence that each probe was successively stronger. This led to the belief that the decreased protein levels were due to the sonicator destroying the sample through denaturation caused by excessive heat. Even though the target protein is thermophilic, that doesn't preclude the possibility that sufficient localised heat was generated to denature it, or that it could be indirectly denatured via host cell protein

degradation. Therefore various levels of sonication were tested, all using 21 micron wavelength as previously used. These were 15 seconds on, 15 seconds off for 1, 3 or 5 cycles, and 15 seconds on, 30 seconds off, also with 1, 3 or 5 cycles. Previously 5 cycles of 15 seconds had been used, with 15 seconds rest in between each cycle.

On the basis of the test, the results of which are shown in Figure 41, it was decided to continue with five cycles of sonication but introduce a 30 second gap between cycles rather than 15. This decision was taken because there appeared to be insufficient sonication with one cycle to lyse the cells and little difference between three or five cycles with regards to a decrease due to over sonication. When comparing the 15 or 30 second rest phase there was a clear reduction in protein in the soluble fraction when a shorter rest phase was used.

As the sample for the test was derived from a single cell pellet to remove variability, there was a smaller volume in each sample than normally would be used. This meant it was likely that these samples were receiving more energy than could be expected in a normal situation. Hence five cycles were chosen on the basis that this would prevent under sonication whilst there being a relatively low risk of over sonication.

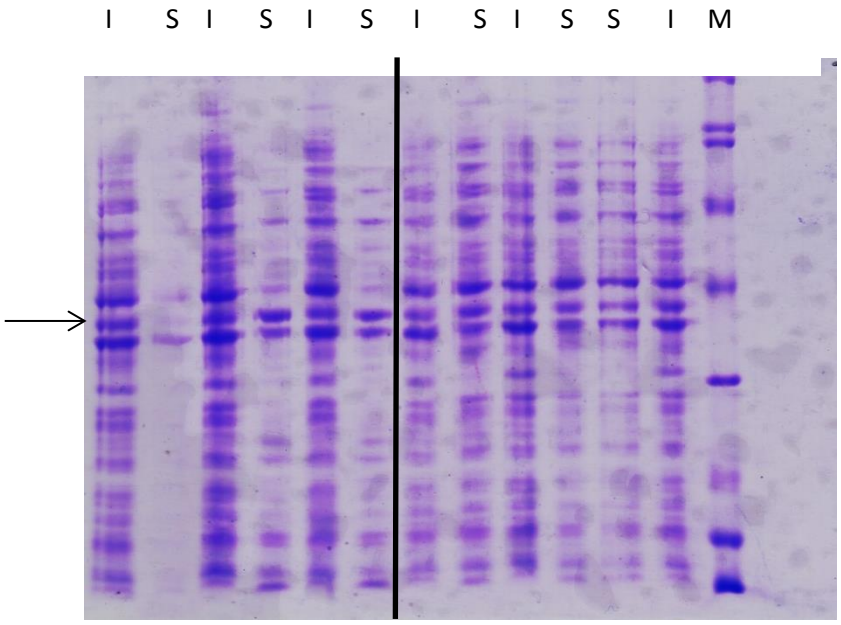


Figure 41 - SDS PAGE gel of Insoluble and Soluble Fractions subjected to sonication at different levels. I = insoluble, S=soluble fractions, M = markers. Arranged in increasing number of cycles, 1, 3 or 5; 15 seconds between cycles to the left of the line, 30 seconds between cycles on the right. Arrow indicates ADH B expression.

2.2.4. ADDITIONAL KINETICS

With sufficient active enzyme now being recovered, further work concentrated upon increasing the substrate concentrations used and working with larger substrates. In addition results were modelled using the Hill Equation, Equation 4 below, in specific cases where negative cooperativity was suspected:

$$v = \frac{V_{max}[S]^h}{K_M + [S]^h}$$

Equation 4 – Generic form of the Hill Equation.

Where h is the Hill coefficient. When h=1, standard Michaelis-Menten kinetics are demonstrated; when h>1, positive cooperativity is indicated and when h<1, negative cooperativity is shown.

As certain Hanes-Woolf plots were non-linear at the left of the plot, tailing downwards from the linear point, this can indicate negative cooperativity because a lower than linear Hanes-Woolf plot indicates a higher than expected rate of reaction for the substrate concentration under study. Therefore, by replotting the same data using the Hill equation and considering the fitness of the data for this equation, it can be seen whether negative cooperativity is present. It should also be stated that this is not the only potential reason for the non-linear Hanes Woolf plots.

One common alternative explanation is an unrecorded background rate as this would have a proportionately larger effect at lower substrate concentrations due to the lower actual rate recorded. Care was taken to ensure adequate controls were done to preclude this possibility, but it is always worth considering nonetheless.

The most intriguing results are those obtained with pentanal (Figure 54 and Figure 55), 2-pentanone (Figure 52 and Figure 53), 3-pentanone (Figure 56 and Figure 57) and 2,3-pentanedione (Figure 58 and Figure 59), as these apparently depart from standard Michaelis-Menten kinetics. Equally surprising is that 3-methylbutanal (Figure 62 and Figure 63) and 1-penten-3-one (Figure 60 and Figure 61) remain firmly demonstrating Michaelis-Menten kinetics. Taken together these results are somewhat equivocal and require further explanation.

The results with 3-methylbutanal could be explained if overall physical bulk of the substrate was irrelevant, with merely the length of the substrate being the key factor. This could also potentially explain 1-penten-3-one as the double bond would shorten the length marginally, but equally the charge of the double bond could be beneficial by inducing a better fit with more charged amino acids in the active site.

In general, it appears that the smallest substrate accepted, ethanal, displays perfect Michaelis-Menten kinetics. Larger substrates demonstrate substrate inhibition, but remain within Michaelis-Menten limits. Further increases in substrate size result in loss of Michaelis-Menten kinetics, and finally any additional size increase results in loss of activity altogether.

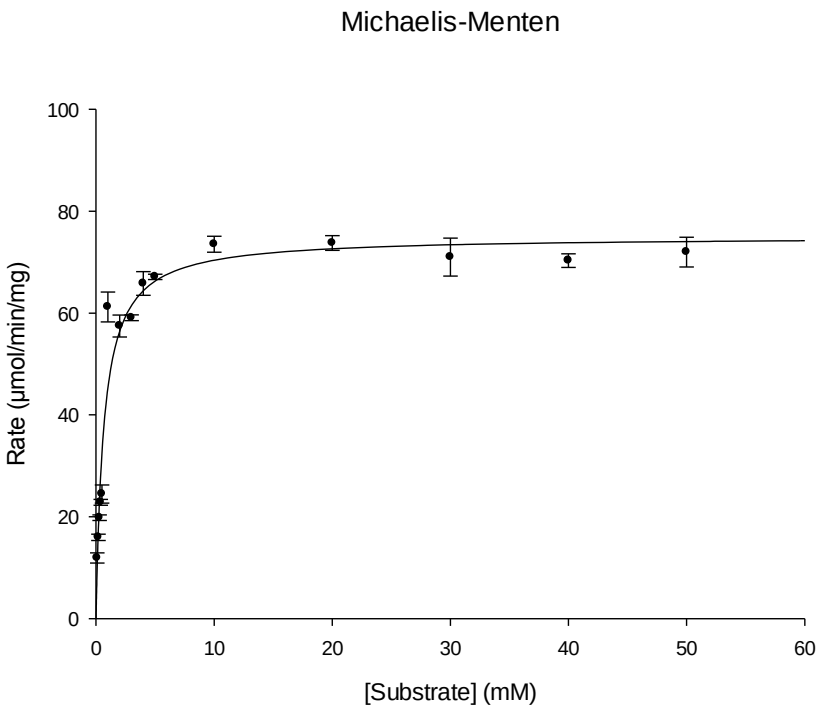


Figure 42 - Graph of rate of reaction vs [S] using ADH B and ethanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

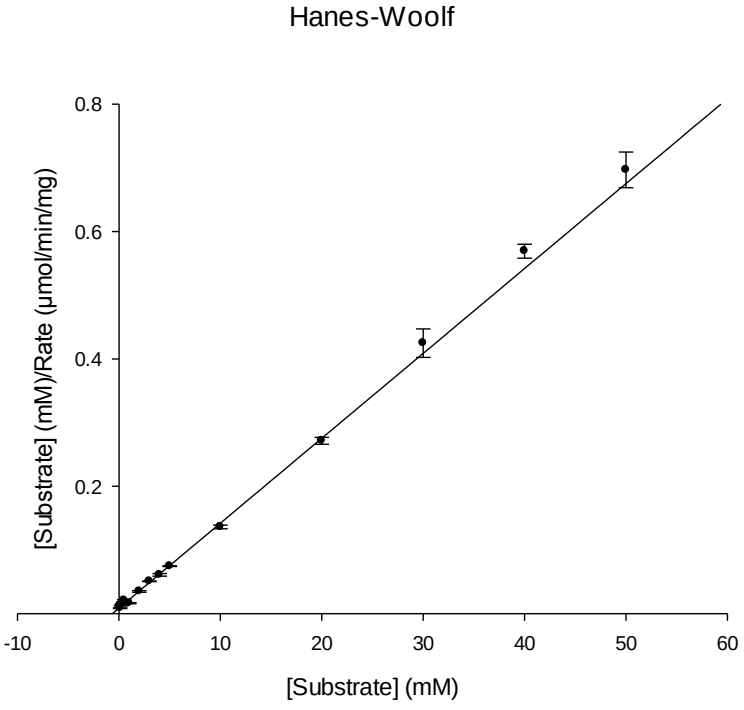


Figure 43 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and ethanal. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

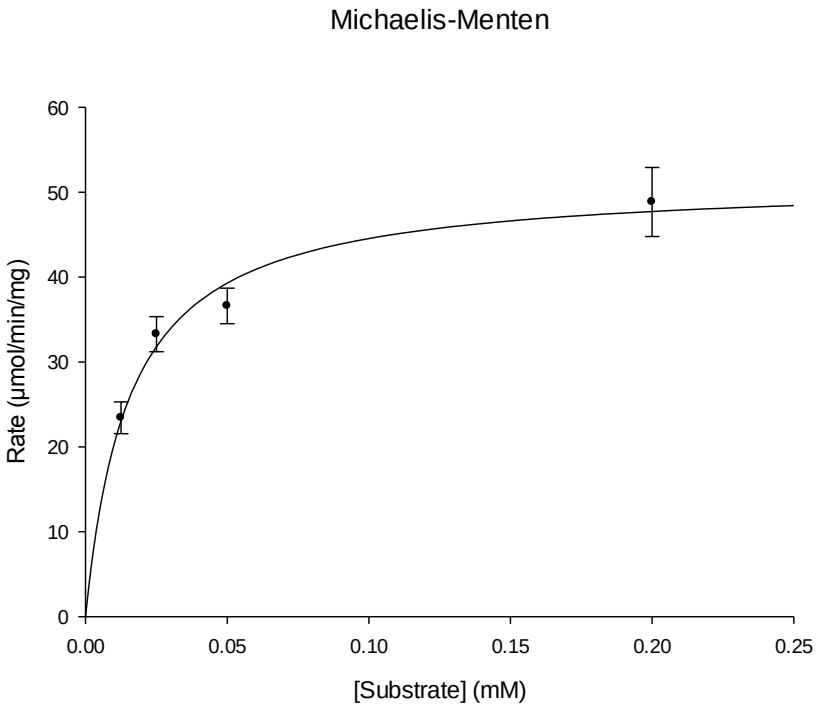


Figure 44- Graph of rate of reaction vs [S] using ADH B and NAD^+. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 10mM 2-butanol, 50 °C, 1ml assay volume.

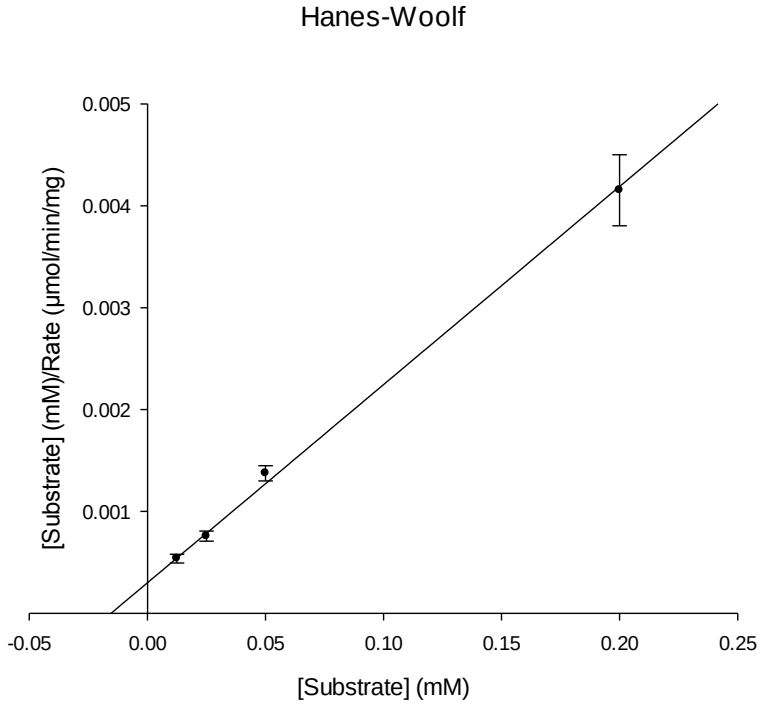


Figure 45 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and NAD^+. Conditions: 50mM citric acid pH 6.0, 10mM 2-butanol, 50 °C, 1ml assay volume.

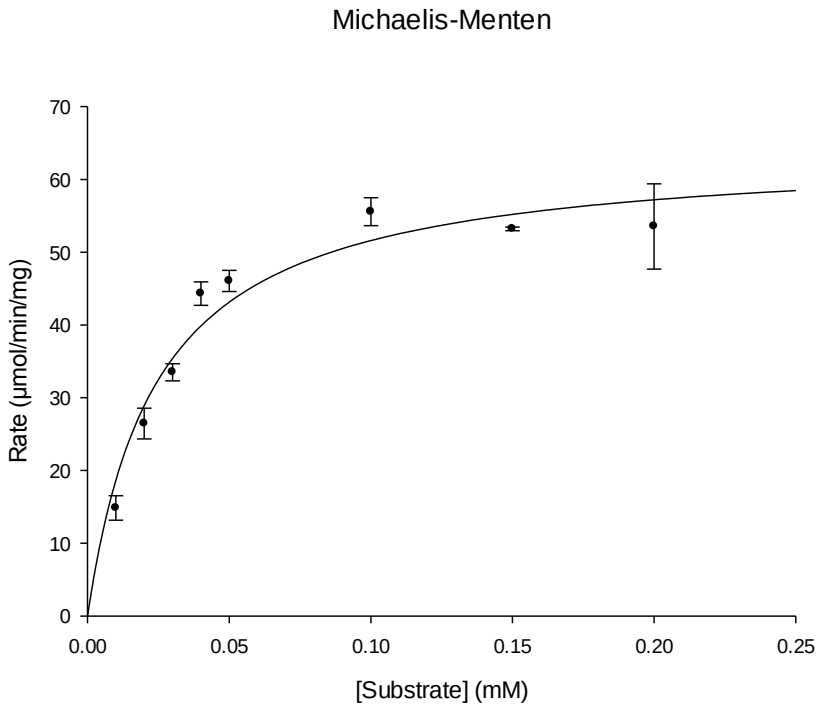


Figure 46 - Graph of rate of reaction vs [S] using ADH B and 2-butanone. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

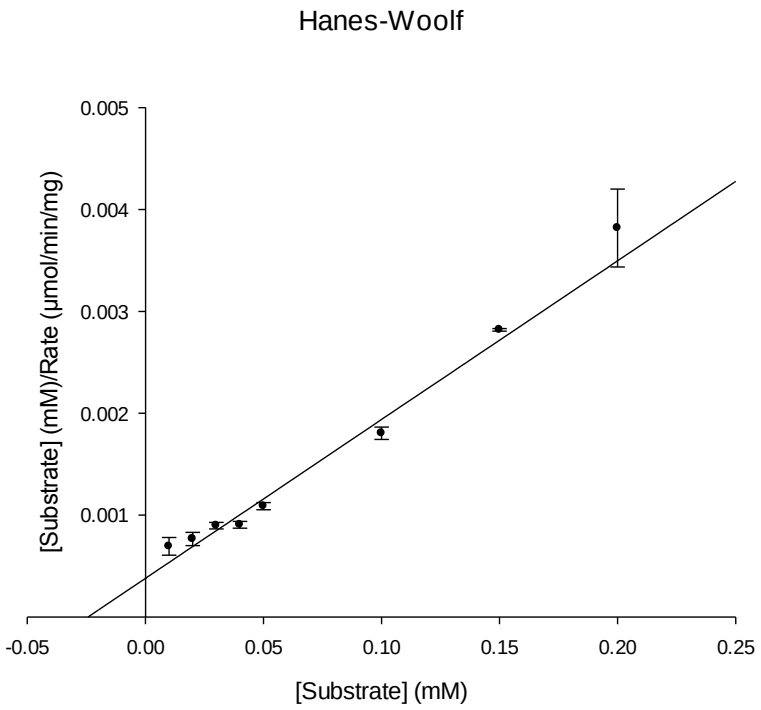


Figure 47 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

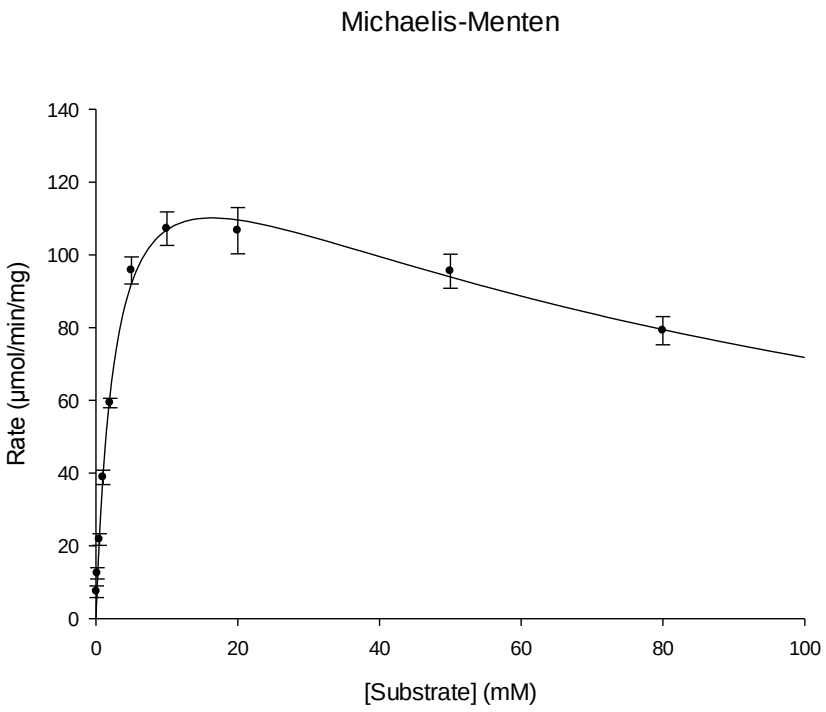


Figure 48 - Graph of rate of reaction vs [S] using ADH B and butanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

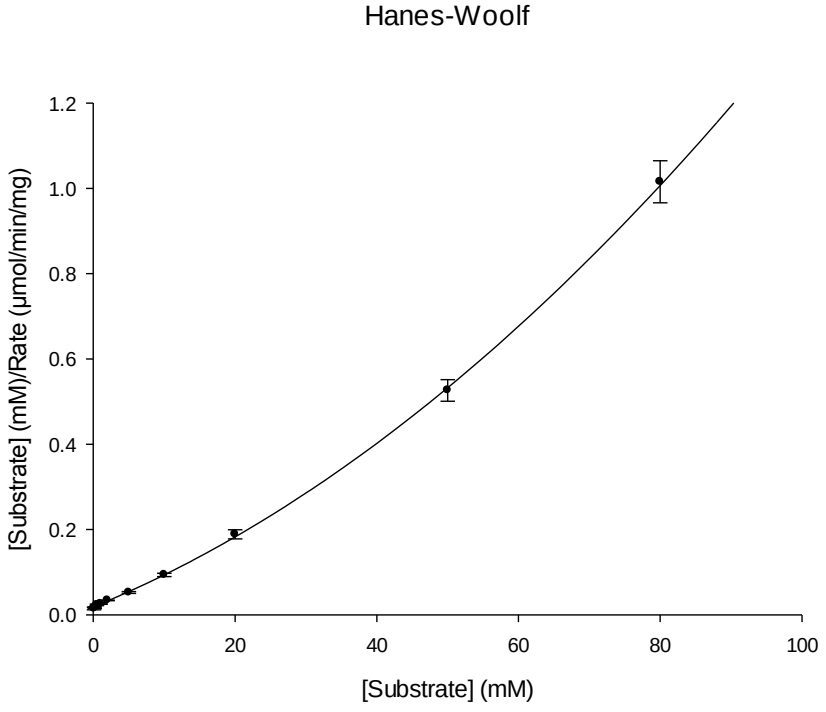


Figure 49 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and butanal, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

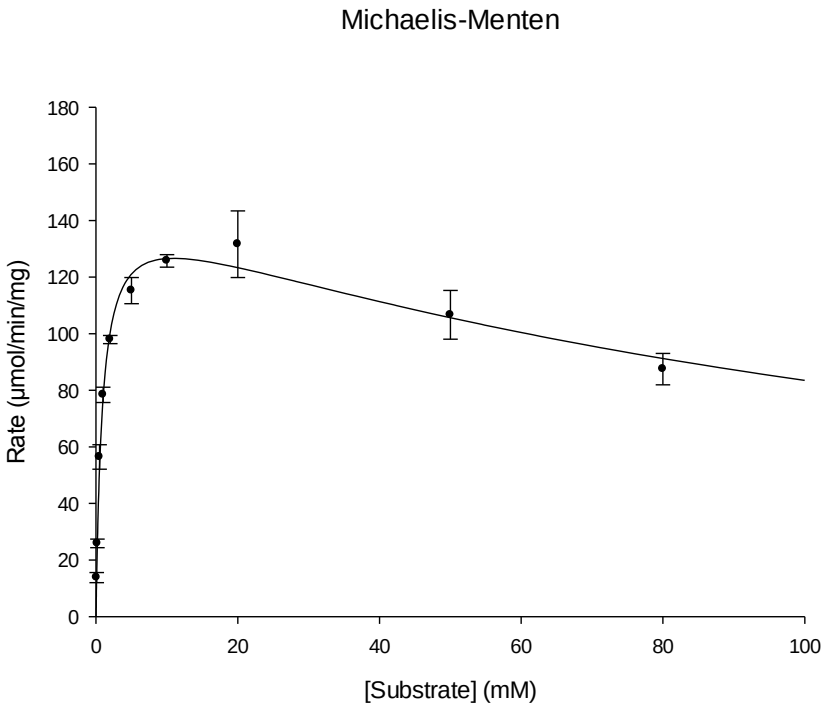


Figure 50 - Graph of rate of reaction vs [S] using ADH B and 2-butanone. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Figure 51 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-butanone, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

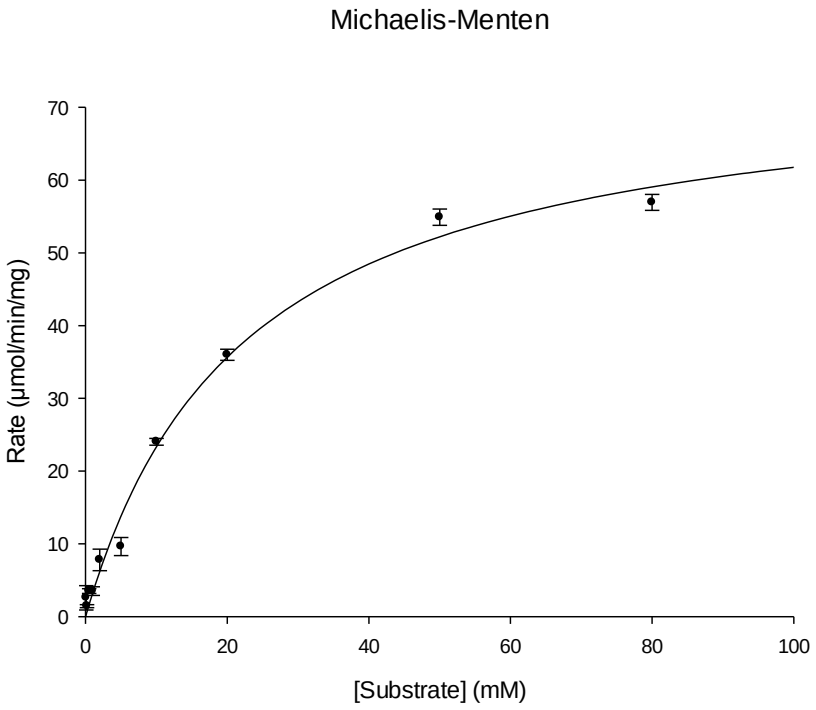


Figure 52 - Graph of rate of reaction vs [S] using ADH B and 2-pentanone. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

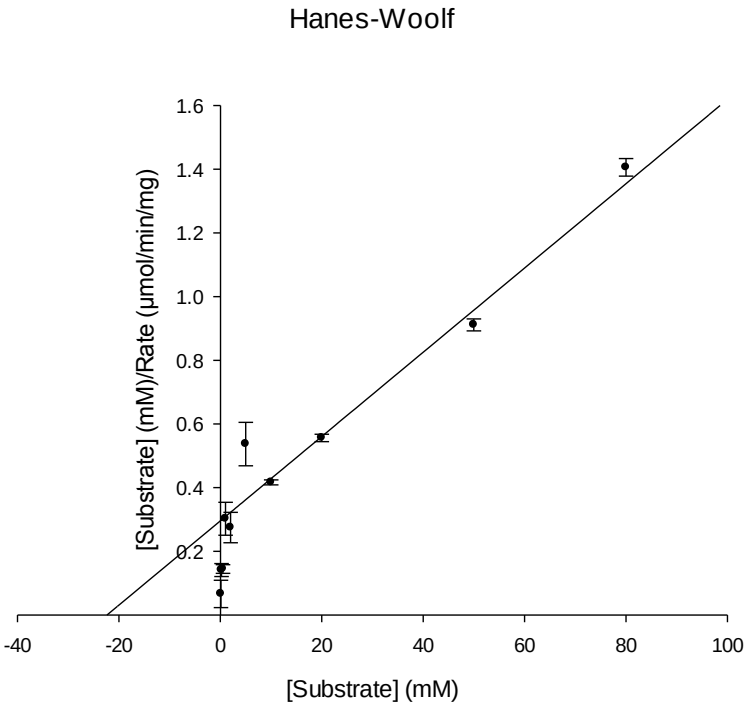


Figure 53 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2-pentanone. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

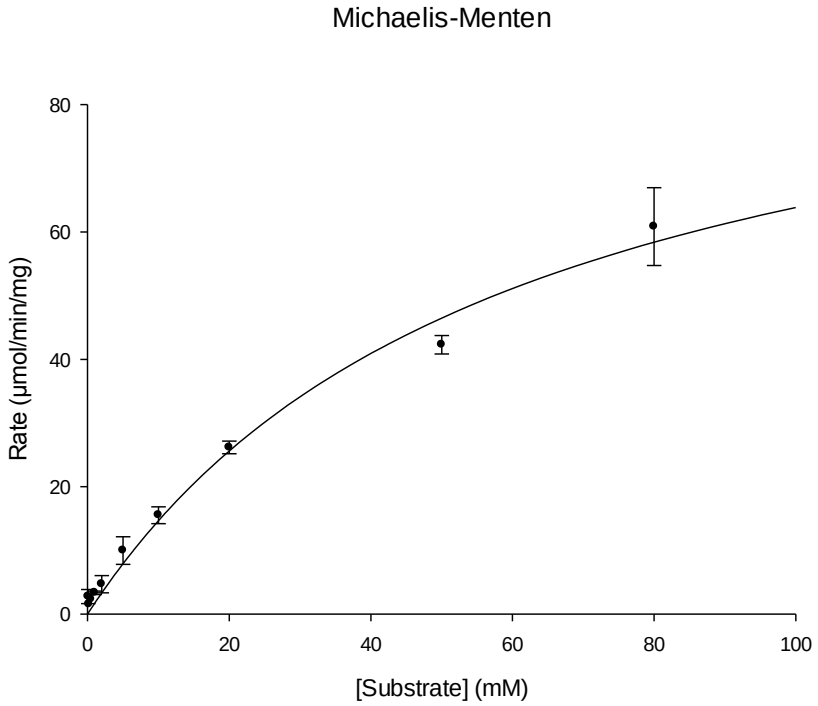


Figure 54 - Graph of rate of reaction vs [S] using ADH B and pentanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

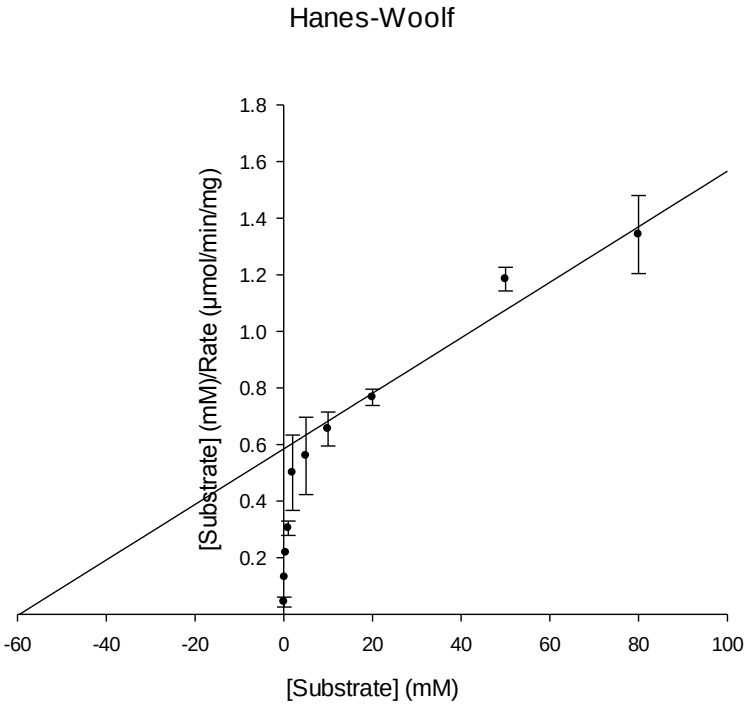


Figure 55 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and pentanal. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

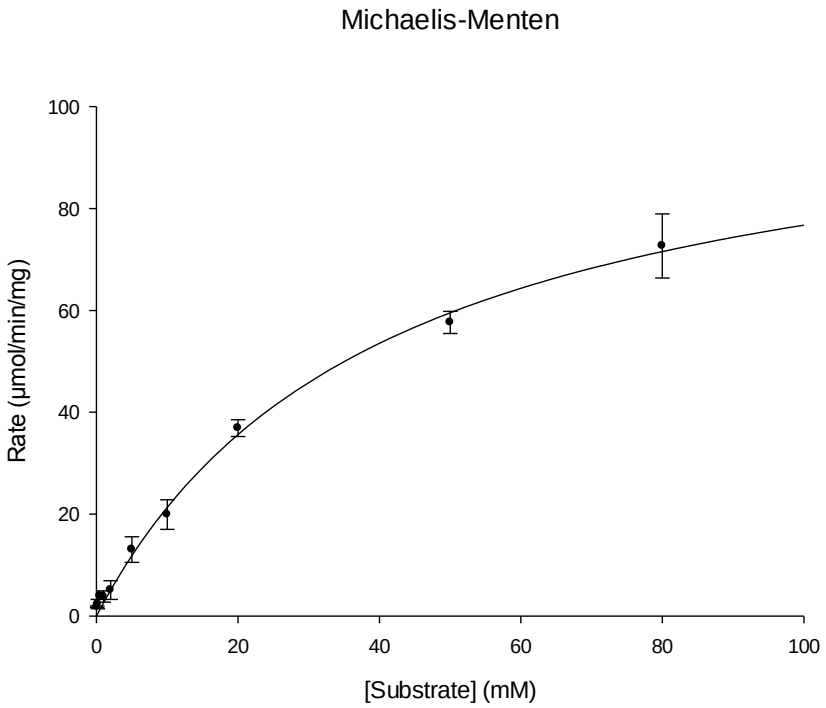


Figure 56 - Graph of rate of reaction vs [S] using ADH B and 3-pentanone. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

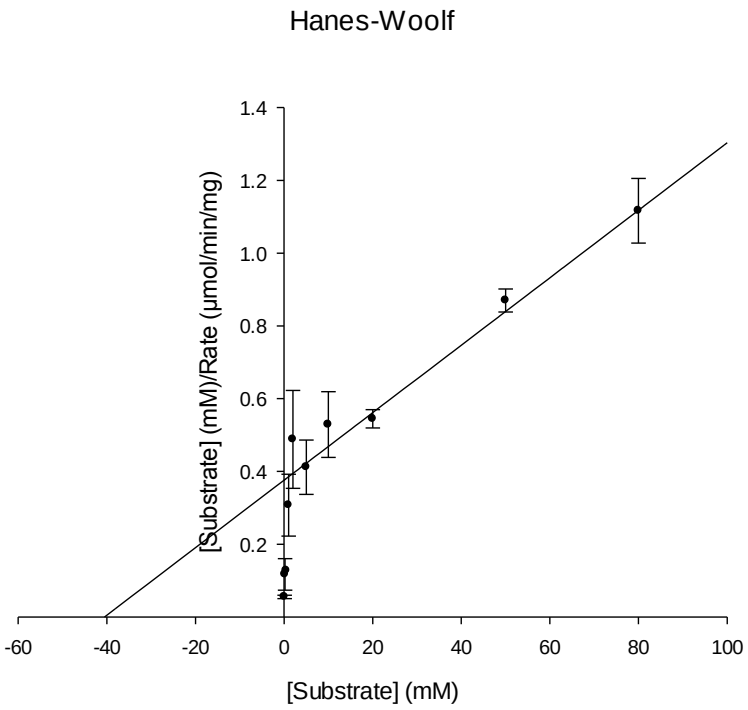


Figure 57 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 3-pentanone. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

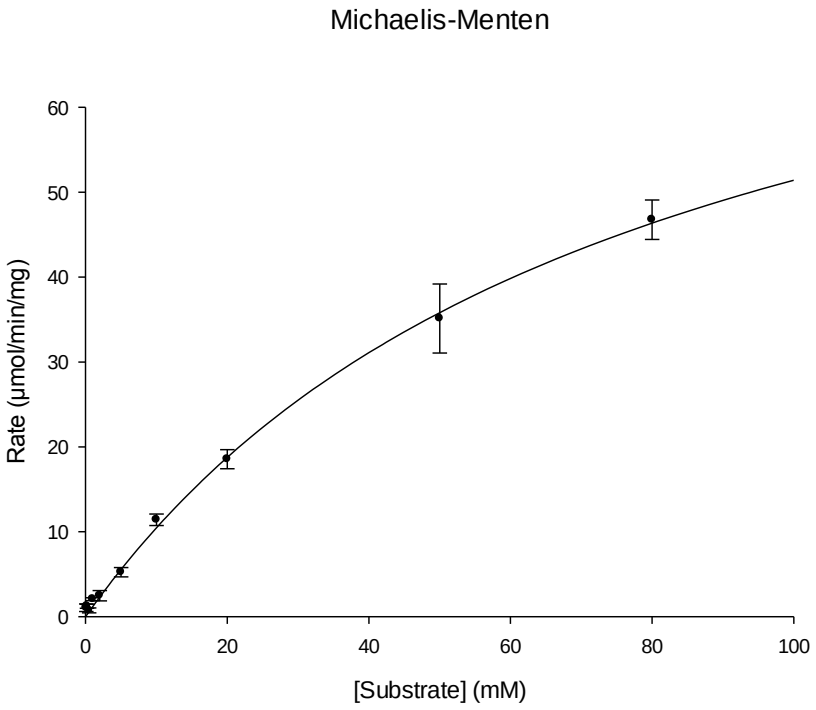


Figure 58 - Graph of rate of reaction vs [S] using ADH B and 2,3-pentanedione. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

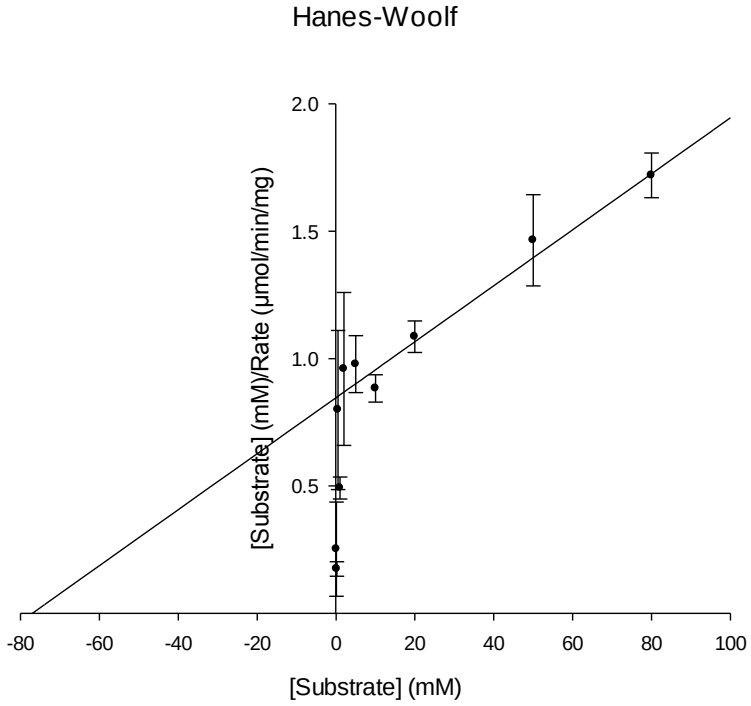


Figure 59 - Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 2,3-pentanedione. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

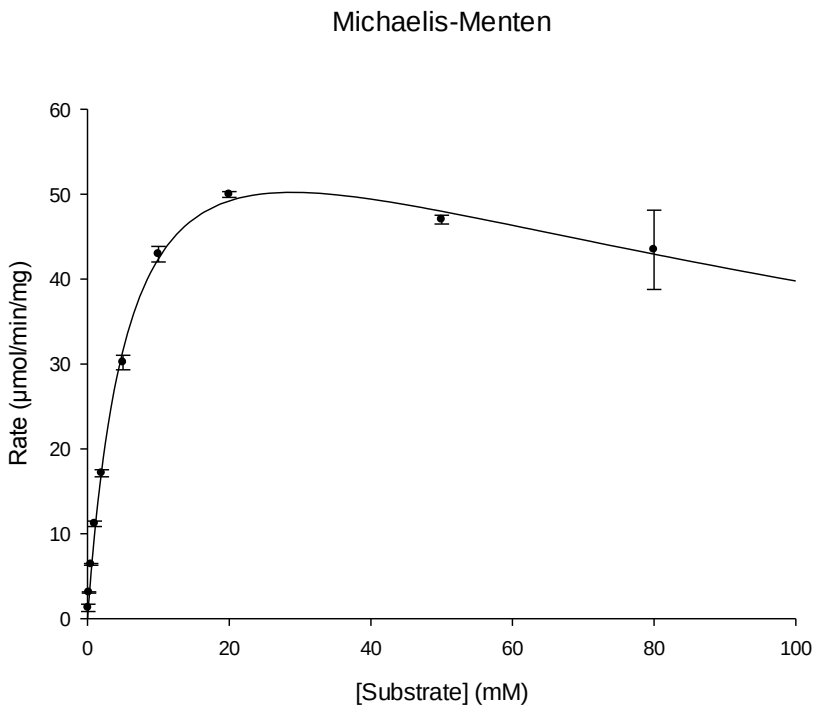


Figure 60 - Graph of rate of reaction vs [S] using ADH B and 1-penten-3-one. Data fitted to a Michaelis-Menten, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

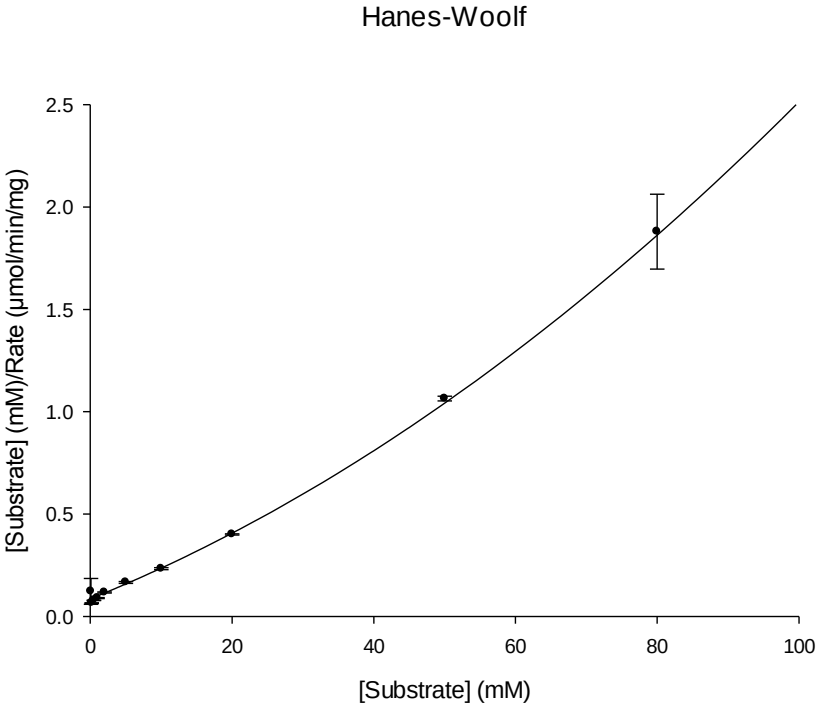


Figure 61 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 1-penten-3-one, data fitted assuming substrate inhibition. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

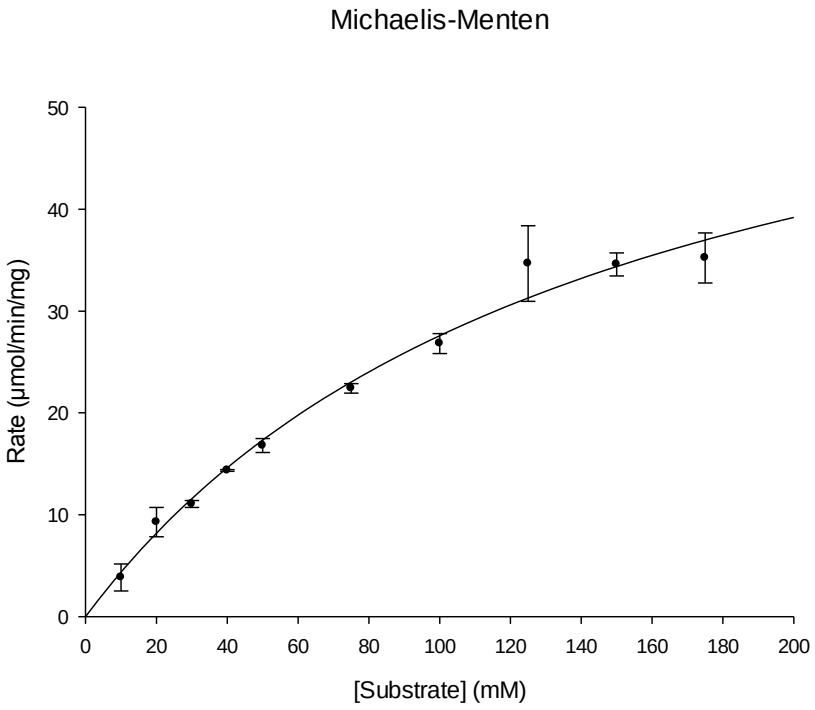


Figure 62 - Graph of rate of reaction vs [S] using ADH B and 3-methylbutanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

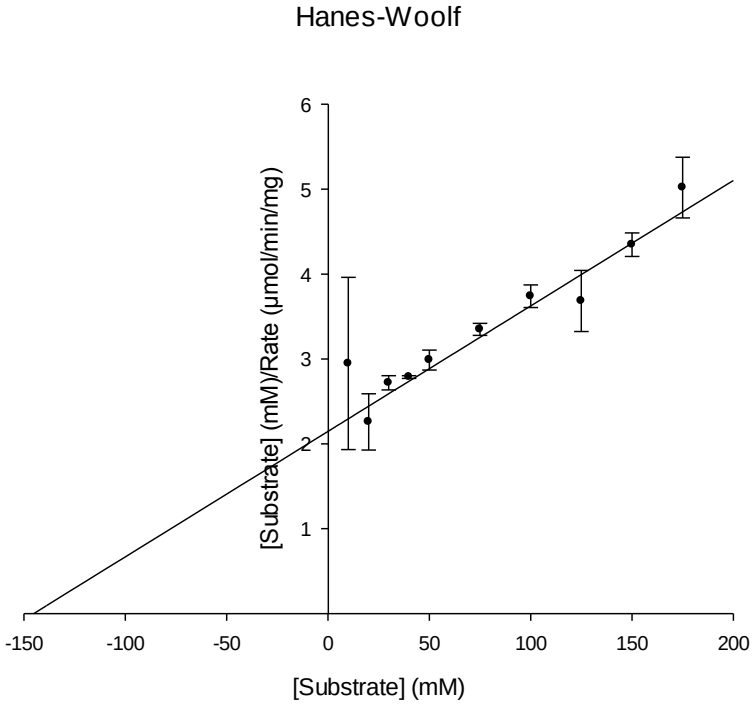


Figure 63 – Hanes-Woolf plot ([S]/v vs [S]) for ADH B and 3-methylbutanal. Conditions: 50mM citric acid pH 6.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})	Michaelis-Menten Kinetics
Ethanal	75.0±1.5	0.66±0.07	-	3000	4500	Yes
NAD^+*	51.4±3.3	0.015±0.004	-	2000	140000	Yes
NADH**	64.2±3.1	0.024±0.004	-	2500	110000	Yes
Butanal	146±7	2.71±0.31	98.4±14.9	5800	2100	Yes, with inhibition
2-Butanone	147±6	0.907±0.119	132±26	5800	6400	Yes, with inhibition
2-Pentanone	75.6±2.8	22.4±2.2	-	3000	130	No
Pentanal	102±13	59.6±14.4	-	4000	70	No
3-Pentanone	108±8	40.6±6.8	-	4300	110	No
2,3-Pentanedione	91.0±10.6	77.1±15.5	-	3600	50	No
1-penten-3-one	71.6±4.8	6.15±0.86	135±30	2800	460	Yes, with inhibition
3-Methylbutanal	67.7±7.5	145±28	-	2700	20	Yes

Table 6 - Further Kinetics Data Summarised. *Cofactor Kinetics, Using 2-butanone as substrate. **Cofactor Kinetics, using 2-butanol as substrate.

Comparing Table 6 with Table 5, particularly the entries in both for 2-butanone, it can be seen that the use of higher concentrations of substrates is justified. Whereas a V_{max} of 73 U mg^{-1} was originally noted, with no inhibition, when investigating the use of higher substrate concentrations, substrate inhibition was noted, and the estimate of V_{max} approximately doubled. Further, the estimate of K_M has trebled. Given the wider range of substrate concentrations used, it is more likely that the second experiment is more representative.

The exception to this is the cofactor kinetics, which appear to be substantially different from what may be expected. Their V_{max} should be more in keeping with the substrate equivalent; this has been attributed to assaying with a large enough concentration of substrates such that inhibition has affected the V_{max}. It is of note that both cofactors have approximately equal kinetic constants though, as at the assay conditions it could be reasonably expected that a bias towards the alcohol forming reaction would be noted due to the mildly acidic pH of the assay.

Whilst there are several cases of deviation from Michaelis-Menten kinetics, the calculated kinetic constants have been retained and these should be considered with caution. Some useful comparisons can still be made though; for example butanal and 2-butanone have identical estimates for V_{max}, yet differ on the K_M, albeit not by a great deal. This slightly increased affinity for the ketone over the aldehyde of equivalent length is mirrored in pentanal/pentanones, with the exception of 1-penten-3-one, which has a substantially lower K_M.

Comparisons are most easily made using the specificity constants, which reflect a three-fold bias towards 2-butanone as opposed to butanal, paralleled with a similar two-fold gap between pentanal and 2-pentanone. A slight difference is also apparent between 2-pentanone and 3-pentanone, with the former being preferred in that case. Perhaps most surprising is the prominently high value for 1-penten-3-one and the prominently low figure for 3-methylbutanal, 460 and 20 mM^{-1} min^{-1} respectively. The considerable difference in activity with C4 and C5 substrates is starkly shown as well, even the high specificity constant of 1-penten-3-one is an order of magnitude lower than 2-butanone.

Due to the above cases where it was suspected that the apparent deviation from Michaelis-Menten kinetics was caused by negative cooperativity, the same data were reanalysed using the Hill Equation as discussed above.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	h	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
2,3-Pentanedione	131(±79)	157(±181)	0.87(±0.14)	5200	30
2-Pentanone	75.4(±5.2)	21.2(±3.9)	0.95(±0.08)	3000	140
3-Pentanone	134(±44)	66.7(±50.0)	0.87(±0.14)	5300	80
Pentanal	933	4383	0.67(±0.14)	37000	10

Table 7 - Summary of supplementary kinetic analysis using the Hill Equation.

From the data in Table 7, it is obvious that aside from pentanal there is no strong case for negative cooperativity with ADH B as h values are approximately 1, within error. Even in the case of pentanal the evidence is somewhat forced, with evidently erroneous kinetic constants. Further work will have to be undertaken to determine the exact reasons for the apparent deviations from Michaelis-Menten kinetics.

As this work was not intended to provide an exhaustive analysis of the kinetic constants and it was not intended to provide a mechanistic analysis, this issue is not catastrophic, and the above estimations are adequate for the purpose required. This is specifically the determination of the substrate scope, and an estimation of the relative affinities of the enzyme for the substrates analysed.

2.2.5. FURTHER WORK WITH ADH B

Stability was another feature of interest of ADH B, with thermostability, solvent stability, substrate inhibition and storage lifetime considered. Due to the thermophilicity of the native species it was naturally assumed that the enzyme would demonstrate similar thermophilicity or better. In addition, it is widely believed that thermophilic enzymes are more likely to demonstrate solvent resistance (Simpson and Cowan, 1992; Cowan, 1997; Vieille et al., 2001). Storage stability would come under a similar heading, as this is meant to encompass long-term storage in chilled conditions; thus thermophilic enzymes should be more stable in these conditions than mesophilic enzymes. Additional causes of degradation whilst stored would be the storage buffer, i.e. chemical stability, or oxygen sensitivity and issues of concentration. Chemical stability is inherently linked with other forms of stability previously mentioned; oxygen sensitivity has little bearing on long term storage (>1 week), as if an enzyme is oxygen sensitive, it will likely inactivate within minutes or hours, comfortably within the timescales of purification (Ma and Adams, 1999; Ying et al., 2009). The concentration of the enzyme was noted but not specifically studied – all enzymes studied were generally eluted in the broad range of 0.1-1.0 mg ml^{-1} as estimated by the Bradford assay. Dilutions were sometimes necessary for effective assays, but no issues were observed with dilutions and for storage stability work only original eluted fractions were studied.

Thermostability was the first aspect studied; 50µl aliquots of ADH B were held at 70 ^{o}C in microcentrifuge tubes to little effect. Samples held for up to one hour had identical rates to a standard sample, but those held for longer were damaged by evaporation, even inside sealed tubes. Subsequent work utilised different methods, using a PCR machine to statically incubate samples as it has a heated lid to prevent evaporation of the samples. This also had the extra benefit of allowing finer control over temperature and the possibility of incubating different samples at a variety of temperatures simultaneously.

Solvent resistance was somewhat more varied. With DMSO, any quantity immediately inactivated ADH B; even 1vol% was enough to negate all activity. With acetonitrile there is little effect upon the addition of 10vol%, with 141 U mg^{-1} activity compared to 153 U mg^{-1} in the absence of solvent, but substantial reductions at higher volumes. Ethyl acetate proved to be somewhat more variable, with an initial drop in activity followed by an apparent rise at higher volumes. As ethyl acetate is only soluble up to 8vol%, larger amounts induce bilayer formation. As an enzyme is likely to be constrained to the aqueous phase, this *de facto* increases the concentration of the enzyme, which leads to an apparently higher activity. Normalised activities indicated that the further addition of ethyl acetate beyond the solubility limit resulted in no change to the activity.

Solvent	**Rate (10vol%) [U mg^{-1}]**	**Rate (20vol%) [U mg^{-1}]**	**Rate (30vol%) [U mg^{-1}]**
DMSO	-	n.d.	n.d.
Ethyl Acetate*	132	154 (136)	177 (138)
Acetonitrile	141	112	53

Table 8 – Summary of solvent resistance data obtained with ADH B and 2-butanone, 153 U mg^{-1} is baseline rate. *Ethyl acetate maximum solubility is 8vol%, thus bilayer formed at higher concentrations. Numbers in brackets refer to normalised activities, assuming enzyme is dispersed solely in the aqueous phase.

Next, using 2-butanone as substrate, increasingly large concentrations of substrate were used to look at the extent of substrate inhibition. Using a baseline of 10mM concentration, 2-butanone was considered at 1M, 2M, 3M and neat (95%) concentrations. In the last case, only the 5µl of enzyme and the 50µl of NADH were added to neat 2-butanone. From a baseline of 146 U mg^{-1} at 10mM 2-butanone, as expected there was a large drop at 1M concentration to 52 U mg^{-1}. This was expected from the estimated K_M of 1mM and K_i of 132mM calculated previously.

More surprisingly, activity was still retained at 2M concentration (38 U mg^{-1}) and even 3M (27 U mg^{-1}). At this point the solubility limit had been reached; therefore no further intermediary stages were tested. Using 95vol% 2-butanone resulted in an unusable assay. The belief was that the NADH was unable to dissolve fully or disperse in the 2-butanone, resulting in no absorbance being detected. The subsequent addition of 200µl buffer to the cuvette allowed for the NADH to be detected and a rate recorded. The rate of 63 U mg^{-1} was substantially higher than expected; it appears that the high level of substrate was interfering with the assay. In particular as an incredibly high absorbance was recorded it is likely that both cofactor and enzyme were dissolved solely into the aqueous phase, resulting in the concentrating effect noted earlier. What was interesting though, is that irrespective of the concentration of 2-butanone, as long as sufficient water was present, activity was maintained throughout. This is an important characteristic for an enzyme to be used industrially, and even though heavy substrate inhibition was present, a high enough rate was still maintained, particularly for a native enzyme.

Due to the frequency of using 2-butanone and ethanal as substrates, and the study of ADH B over a period of months using the same sample, it was possible to look retrospectively at the storage stability of the enzyme. Typically one or other of the standard substrates was used as a type of control to ensure activity remained over the long term. As this was a retrospective analysis, there were relatively few points to analyse, and in the case of 2-butanone no degradation was noted due to the scatter of the data available.

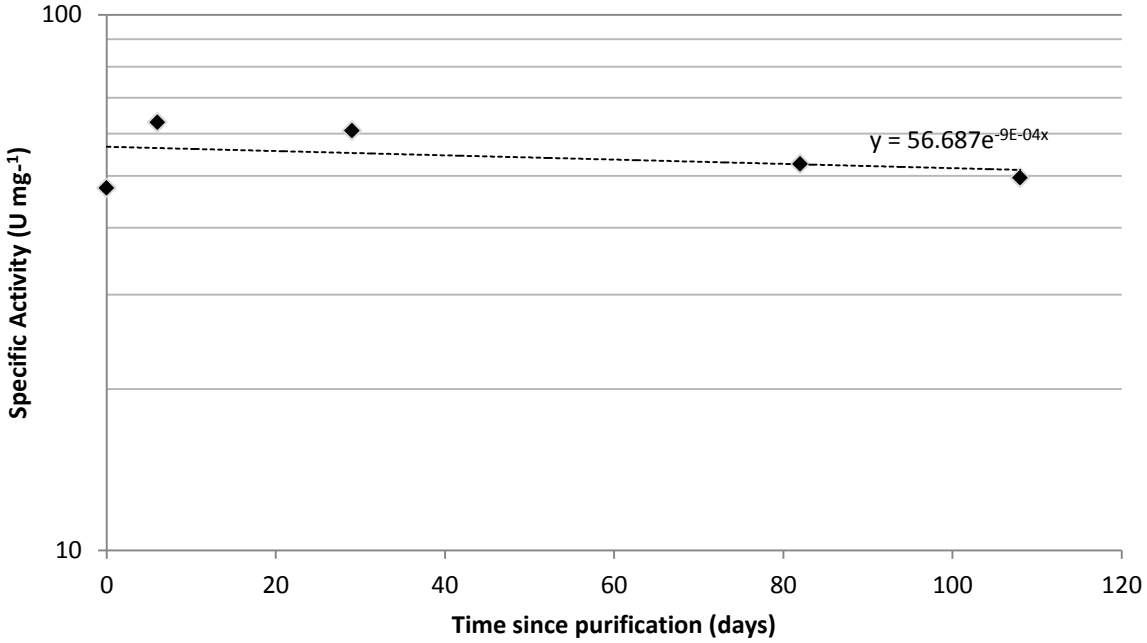


Figure 64 – Graph of specific activities of ADH B recorded with ethanal in assays conducted after different periods of static incubation at 4 °C. Assay conditions: 50mM citric acid pH 6.0, 50 °C, 0.2mM NADH, 1ml volume assay.

Given the estimate of the rate of degradation provided in Table 9, it is possible to determine an estimate for the half-life of the enzyme in storage conditions. The definition of the half-life is given by Equation 5 below:

$$t_{1/2} = \frac{ln(2)}{\lambda}$$

Equation 5 - Definition of Half-Life

Where λ is the decay constant.

$$N_t = N_0 e^{-\lambda t}$$

Equation 6 - Exponential Decay Formula

Where N_t is the current activity of the enzyme, N_0 is the initial rate of the enzyme and t is the number of days since purification, in this specific case.

As Figure 64 provides an estimate of the gradient, or the decay constant, of 0.0009, which can be directly used in Equation 5 to estimate the half-life:

$$t_{1/2} = \frac{ln(2)}{0.0009}$$

$$t_{1/2} = 770$$

770 days is obviously only an estimate, but considering that the 2-butanone results were unable to give a decay rate of any kind, it was likely that any half-life was likely to be relatively high. In terms of stability, it can be stated unequivocally that ADH B is extremely stable in the elution buffer and stored at 4 °C. As this buffer contains substantial amounts of imidazole,

which can have a stabilising effect on enzymes, this could be a major factor. Other possibilities are the high salt content of the buffer, or the pH and temperature alone. As detailed studies were not carried out on long term storage stability, it isn't possible to state if the stability is inherent or dependent on the buffer. That said, it would not be unexpected if a thermophilic enzyme were to be stable over the long term whilst chilled, as long as the storage buffer was not a net negative on stability. Additionally, no activity was lost when ADH B was inadvertently stored at room temperature overnight.

Lastly, alternative substrates were also screened. In most cases, these were additional aldehydes and ketones, but two specific cases were also tested, L-serine and L-threonine. This was due to the annotation of ADH B as a Threonine Dehydrogenase. A summary of these results is below.

Substrate	Activity
L-serine	No
L-threonine	No
2,4-Dimethyl-3-pentanone	Yes
Isobutanal	Yes
Crotanal	Yes
Methanal	No
Hexanal	No
2-Hexanone	No
6-Methyl-5-hepten-3-one	No
Ethyl-2-oxo-4-phenylbutyrate	No
4-Phenyl-2-butanone	No
Ethyl 4,4,4-trifluoroacetoacetate	Yes

Table 9 - Summary of secondary screen of substrates; substrates tested at concentrations from 10-100mM.

2.2.6. SEQUENCE ALIGNMENT

For this multiple sequence alignment it was possible to include both closely related sequences from the NCBI database and the most closely related sequences held by the PDB. Ten of each were selected and aligned together.

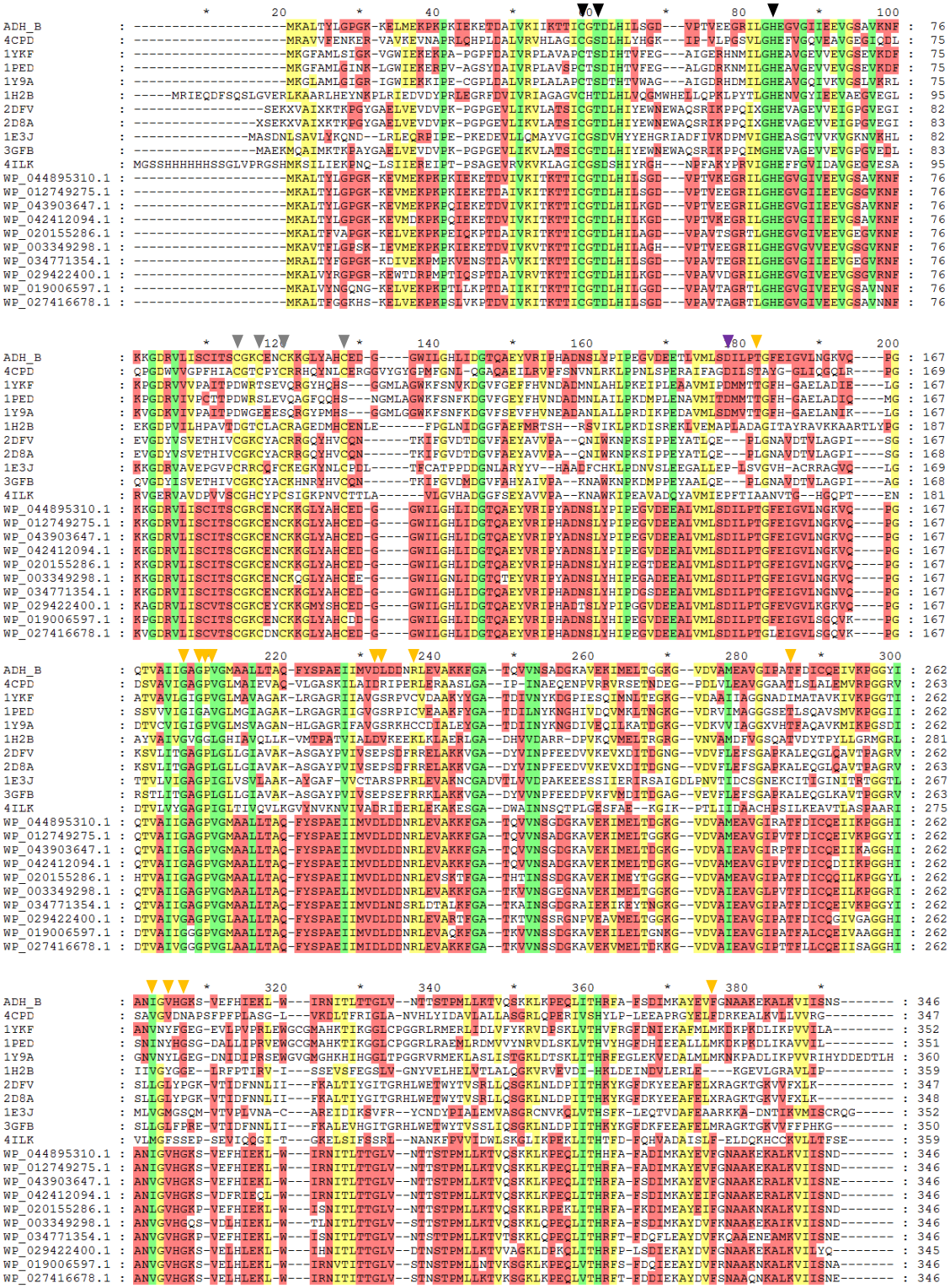


Figure 65 - Multiple Sequence Alignment of ADH B with most similar sequences associated with crystal structures and most similar sequences from the NCBI RefSeq database. Orange triangles indicate NAD(P) binding site. Black triangles indicate catalytic Zn binding site. Grey triangles indicate structural Zn binding site. Purple triangle indicates a dual NAD(P) and catalytic Zn binding site. All these sites provided via the NCBI Conserved Domain Database and obtained through the Conserved Domain Search. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to crystal structures from the PDB or proteins from the NCBI RefSeq Database. Proteins are all annotated as ADHs from genera *Bacillus, Geobacillus, Marinobacterium, Cohnella* and *Alicyclobacillus*. The crystal structures are annotated as follows: 4CPD is an ADH from *Thermus sp. ATN1*, 1YKF is a $NADP^+$ dependent ADH from *Thermoanaerobium brockii*, 1PED is a secondary ADH from *Clostridium beijerinki*, 1Y9A is an ADH from *Entamoeba histolotica*, 1H2B is an ADH from *Aeropyrum pernix*, 2DFV is a Threonine Dehydrogenase (TDH) from *Pyrococcus horikoshii*, 2D8A is a putative TDH from *Pyrococcus horikoshii*, 1E3J is a Sorbitol Dehydrogenase from *Bemisia tabaci*, 3GFB is a TDH from *Thermococcus kodakarensis*, 4ILK is an ADH from *E. coli.* Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997).

Though there are few residues absolutely conserved across sequences with such little homology, it is noticeable that all three residues noted as part of the catalytic zinc binding site are all maintained, along with an appreciable number of adjacent residues. Given that most but not all of the sequences have been annotated as ADHs, it might reasonably be assumed that the structural rather than the catalytic zinc might be conserved because ADHs, TDHs and sorbitol DHs are all in the same protein family, the medium chain dehydrogenase superfamily (Hedlund et al., 2010). This is not the case though; the residues associated with the structural zinc are moderately well conserved, but not in the case of three ADHs.

Some of the residues associated with the NAD(P)H binding site are even less well conserved, with the least similar sequences from the PDB differing in approximately half of these residues. It should also be noted that the sequences from the NCBI database are very similar to ADH B and only differ in one case at one position – G307K.

Examining the top five sequences from the PDB may provide more information about ADH B. 4CPD is a solvent-resistant ADH from *Thermus sp. ATN1* that exclusively produces *(S)*-enantiomer alcohols. It is NADH dependent and has also demonstrated dismutase activity. Its substrate scope is wide and includes primary and secondary aliphatic alcohols, cyclic alcohols and aromatic alcohols, as well as their aldehyde and ketone equivalents. In the reductive direction, high activity was constrained to aldehydes and cyclic ketones. It is also a homotetramer (Höllrigl et al., 2008). The structural zinc is co-ordinated to several residues - C89, C92, C95 and C103 as numbered originally, corresponding to C114, C117, C120, C128 in Figure 65. The catalytic zinc is co-ordinated to C38 (59), H59 (85), D152 (178) and a water molecule, which are as above, but differ with regards to S40 (61) which is shown in Figure 65. This residue is instead the presumed proton donor rather than necessarily co-ordinated to the zinc ion, but nevertheless remains crucial to catalysis. The remainder of the active site is hydrophobic and consists of residues I49 (73), F85 (110), V108 (133), F115 (140), V268 (305), L282 (320), V283 (324) L292 (333), and potentially Y111 (136) with the hydrophobic nature of the active site lending itself well to binding hydrophobic substrates. H43 (64), D200 (232), R201 (233), R205 (237), V244 (281) and V260 (300) are key for cofactor binding (Man et al., 2014a). The active site is not annotated on the MSA, but the putative cofactor binding residues only partially match.

1YKF is a tetrameric NADPH-dependent ADH from *Thermoanaerobium brockii*, one of the first crystal structures to be obtained from prokaryotic ADHs, and also one of the first NADPH dependent dehydrogenases. Cofactor specificity is dependent on positions GSR198-200 and Y218 by the original sequence numbering, corresponding to positions 232-234 and 254 in

Figure 65. These residues bind the phosphate of the NADPH, which is prohibited in 4CPD due to the presence of D200 (232) which is a negatively-charged residue and binds to NADH instead, and I202 (234) which is hydrophobic and the equivalent residue to S199 in this enzyme (Korkhin et al., 1998; Man et al., 2014a). These positions are not all noted as being involved in cofactor binding, according to the NCBI database, but are conserved among three of the ADHs. Catalytic activity is co-ordinated by residues Q60 (85) and S39 (63), again partially disagreeing with the notation above. D150 (179) is also believed to be critical to determining the enantiospecificity of the enzyme; in this case no preference is predicted (Korkhin et al., 1998; Radianingtyas and Wright, 2003). When used to catalyse the production of 2,3-butanediol from *(R)*-acetoin, the enzyme exclusively produced *(R,R)*-2,3-butanediol (Yan et al., 2009). Elsewhere the situation is more complex, with the concentration of 2-propanol (the sacrificial substrate) affecting the kinetics in terms of the enantiomeric excess over the course of the reaction but ultimately a racemic mixture is produced, as predicted. It is also possible to affect the enantiomeric excess by altering the temperature of the reaction or using solvents (Yang et al., 1997).

1PED is a tetrameric NADPH-dependent ADH from *Clostridium beijerinckii*. Similar to 1YKF in many ways, including sequence identity, structure and history, its crystal structure was determined at the same time as 1YKF and they were revealed to have very similar 3D structures and 75% sequence identity. In this case the catalytic zinc is co-ordinated by C37 (59), H59 (84), D150 (179) and Q60 (85), although the latter residue retracts from the zinc upon the addition of the co-factor. As previously, D150 (179) is key for determining the enantiospecificity of the enzyme and again no preference is predicted from the orientation. All other structural features of 1PED are identical to 1YKF (Korkhin et al., 1998). 1PED has a relatively constrained substrate scope, but includes primary and secondary alcohols and their respective aldehydes and ketones; the largest substrate tested was 2-pentanone. The largest turnover numbers were obtained with aldehydes; however, the specificity constants were larger with the ketones, with ethanal and acetone the most preferred, respectively. Under the conditions tested there was little activity with corresponding alcohols (Ismaiel et al., 1993).

1Y9A is a tetrameric $NADP^+$ dependent ADH from *Entamoeba histolytica*, the only eukaryote of the five enzymes. It has activity with isopropanol, ethanol, acetone and ethanal, with its presumed purpose being in ethanol production from glucose in anaerobic conditions (Kumar et al., 1992). Subsequent work attempted to determine if there was a specific cause for the relative thermolability of 1Y9A compared to 1YKF, despite their high sequence identity. A single residue was significant in this regard, with P275D (314) substantially increasing the thermostability of 1Y9A and the reverse significantly reducing the thermostability of 1YKF. This residue is found in the dimerization interface and is believed that the proline residue is able to increase the hydrophobic interactions (Goihberg et al., 2008).

Lastly, the only representative of the *Archaea* in the top 5 is 1H2B. This is an NADH-dependent tetrameric ADH from *Aeropyrum pernix* with a broad substrate scope of alcohols, aldehydes and ketones, but with preference for cyclic substrates. With respect to alcohols it is most active with C4-C5 primary and secondary alcohols and large cyclic alcohols such as cycloheptanol and cyclooctanol. Particularly important residues noted include D218 (232), which is thought to determine the cofactor specificity. It is highly conserved in NADH-dependent ADHs, but replaced with glycine in NADPH-dependent enzymes. Other key residues for cofactor binding include R354 (355), V262 (281), V219 (233) and H55 (60). The catalytic zinc ion is co-ordinated to C54 (59), H79 (84) and D168 (182) with an additional residue, E80 (85) binding in the absence of bound inhibitor or substrate. The structural zinc ion is co-ordinated with C112 (117), C115 (120) , C123 (128) and D109 (114), but is only present part of

the time; in the absence of the ion, C123 and C115 form a disulphide bond (Guy et al., 2003; Littlechild, 2011).

Overall, it is not surprising that the annotation with respect to the active site, metal ion co-ordination and cofactor binding is not entirely reliable – this is to be expected of all automatic annotations. That said, the majority of noted residues are believed to be involved in the areas annotated, so it is not entirely wide of the mark. Highlights of the above information include the potential to raise the T_m of ADH B by altering a single residue, in this case H314P. Further, having a substrate scope limited by size would seem to be in keeping with the above examples, as would the cofactor specificity, with D232 indicating a preference for NADH. The enantiospecificity of ADH B has not been considered experimentally; however, should further work include crystallisation, then residue D179 is one that could indicate any preference. That said, if similar enzymes are considered, then ADH B likely has little to no enantiospecificity.

2.2.7. SUMMARY

ADH B is an alcohol dehydrogenase limited to five carbon length substrates only. Given this restriction, with the sole exception of glycerol, all substrates were accepted, including halogenated, branched and functionally diverse aldehydes and ketones.

With substantial solvent tolerance and resistance, it is a prime candidate to consider for industrial use, particularly with regards to ethyl 4-chloroacetoacetate, an intermediate in the production of statins.

High thermostability is another feature of this enzyme, with no irreversible denaturation noted after one hour incubation at 70 ^{o}C. This is a particular area that should be considered further, as all experiments were conducted at 50 ^{o}C; therefore rates may be more favourable at higher temperatures.

Whilst the foundations of extreme storage stability have not been determined, it is clear that with an estimated half-life of 770 days, ADH B can be stored for very lengthy periods of time without resorting to methods such as lyophilisation.

Finally, as ADHs generally have a rather large bias towards aldehydes or ketones as substrates, typically at least an order of magnitude variance in specificity constant, it is surprising that only a maximum threefold difference was recorded with ADH B and substrate size was of sole concern.

2.3. ADH C

ADH C was previously under investigation as an acetoin reductase, to be used as part of the same metabolic pathway indicated in Figure 27. Using the same assay conditions with ADH C as with previous enzymes resulted in no detectable activity, and necessitated changing buffer from citric acid (pH 6.0) to sodium phosphate (pH 7.0). This was attributed to citric acid's ability to chelate metal ions rather than the pH change. As ADH C is annotated as a zinc related ADH, it is certainly plausible that citric acid could chelate the zinc required for activity.

Previous work had confirmed activity with acetoin, and therefore the full substrate scope was sought. Looking at similar substrates in terms of size and functional groups, activity was additionally seen with 2,3-butanedione and 2,3-pentanedione. As the diketones had activity, the equivalent monoketones were considered; however, no activity was noted with 2-butanone, acetone, butanal, 3-pentanone or the less similar 2,4-dimethyl-3-pentanone and 4-phenyl-2-butanone.

No further substrates were available for testing, and due to the limited substrate scope no further work was conducted with this enzyme. Its lack of activity in citric acid buffer prompted a full change to sodium phosphate in subsequent assays with other enzymes, to preclude the possibility that other metal-dependent enzymes may falsely appear to be inactive.

3. INVESTIGATIONS WITH NOVEL ENZYMES

All genes referred to in this section were identified, isolated and cloned into *E. coli* BL21 (DE3), using pET28a to provide an N-terminal hexa-histidine tag, as part of the investigation. They were selected on the basis of their annotation and similarity to alcohol dehydrogenases with industrially relevant substrate profiles. As relying on such annotation is imprecise at best, it is necessary to approach each enzyme with an open mind and determine its substrate scope independently.

Due to the potential imprecision of the annotation and the potential for cloning of the relevant genes to fail, the initial list of putative ADHs was drawn up with very loose boundaries. Whilst it was accepted that this may include some false positives, it was believed that this would minimise false negative results. This initial list comprised a total of 61 potential targets which was eventually reduced to an initial batch of six genes. After only two of these genes were successfully cloned into *E. coli*, a further batch of ten genes was added. Of the sixteen putative ADHs targeted, a total of 14 have been successfully cloned and expressed, but due to time constraints fewer have been assayed and characterised, the details of which are contained in this chapter.

ADHs were chosen for cloning on the basis of annotation and the likelihood that the substrates accepted by the enzymes would be of industrial interest. Therefore enzymes annotated as being aryl-alcohol dehydrogenases were given higher priority than those believed to accept small aldehydes. Enzymes conserved in related species and genera and consistently annotated as accepting larger substrates were also prioritised. Finally, enzymes annotated as being Aldo-Keto Reductases or part of the Medium Chain Dehydrogenase/Reductase family were prioritised over Short Chain Dehydrogenase/Reductase and iron-containing ADHs.

3.1. ADH D

The gene coding for ADH D was initially isolated in the first batch of targets, and ADH D was one of two proteins to be successfully produced recombinantly from that batch. Figure 66 is an SDS-PAGE gel of ADH D post IMAC, indicating successful protein production and purification.

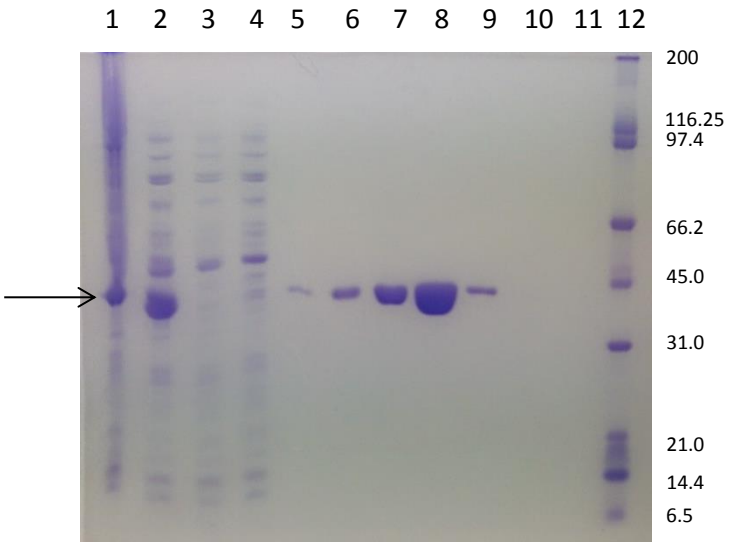


Figure 66 - SDS PAGE gel of ADH D. Lane 1: insoluble fraction, lane 2: soluble fraction, lane 3: 2nd flowthrough, lane 4: 0% His-Elute Buffer, lane 5: 1%, lane 6: 2.5%, lane 7: 5%, lane 8: 10%, lane 9: 30%, lane 10: 100% His-Elute Buffer, lane 12: Protein markers. The numbers on the right indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 38.0 kDa

Despite early testing indicating activity with several substrates, further work was deferred in favour of ADH B because of the latter's preference for NADH as cofactor rather than NADPH. As refinement of the experimental procedure was required, it was preferred to do this with an NADH utilising enzyme rather than NADPH for cost considerations. The preliminary assay data are summarised in Table 10 below:

Substrate	Activity
2-Butanone	No
Ethyl 4-chloroacetoacetate	Yes
Ethyl 2-oxo-4-phenylbutyrate	Yes
4-phenyl-2-butanone	No
Ethyl 4,4,4-trifluoroacetoacetate	No
Cyclohexanone	No
Cycloheptanone	No
(R)-Carvone	No
Acetone	No
Hexanal	Yes
Propanal	Yes
6-Methyl-5-hepten-2-one	No
Methanal	Yes
Acetoin	No
Octanal	Yes
Decanal	Yes
Dodecanal	Yes
Butanal	Yes
Propiophenone	No
Ethyl pyruvate	Yes
Ethyl acetoacetate	No

2,4-Dimethyl-3-pentanone	No
Isobutanal	Yes
3-Hydroxy-3-methyl-2-pentanone	No
3,4-Hexanedione	Yes
Crotanal	Yes
Furfural	Yes
5-Norbornene-2-carboxaldehyde	Yes
2-Decanone	No
2,2,2-Trifluoroacetophenone	No
Acetophenone	No
6 Methyl-5-hepten-2-one	Yes

Table 10 - Initial screen of ADH D. Conditions: 0.2mM NADPH as cofactor, 50mM Sodium Phosphate, pH 7.0, 1ml assay volume, 50 °C.

This wide ranging initial screen formed the basis of an enhanced screening protocol, based around primary and secondary substrate screening. A standard group of substrates were compiled, incorporating linear, branched, cyclic and aromatic aldehydes, ketones, diketones, alpha, beta and gamma keto-esters, their acid equivalents, halogenated substrates and industrially relevant examples of the above. The range of substrates included in the primary substrate screen was deliberately chosen in stark contrast to previous methods that utilised a very limited initial screen. The range and varied complexity of the substrates, whilst keeping to a relatively limited number of substrates, allowed for a comprehensive opening salvo of substrates to give the maximal chances of determining whether an enzyme demonstrated activity whilst reducing the volume of assays to a practicable level. Given that any specific enzyme annotated as an alcohol dehydrogenase may not in reality be an ADH, may have particular preferences with regards to substrates or assay conditions, may not express or purify in an active form, or may be unstable, the effectiveness of a screening protocol is paramount to the success of the research. Specific substrates were chosen based upon their prevalence in literature, ease of acquisition, ease of assay and lack of superfluous functionality in addition to their variety. Therefore the standard small aldehyde was butanal rather than ethanal as used previously due to the latter's high evaporation rate, difficulty in creating accurate stock solutions and potential for degradation. A small number of previously studied industrially relevant substrates were retained as well; all those that were easily assayed were kept.

Based upon the results of the primary screen, additional substrates were considered on a case-by-case basis for each enzyme with regards to activities detected and availability of substrates. Typically this would be done in order to delineate the exact substrates that could be accepted by an enzyme and those that would not. Generally, however, this proved to be more difficult than anticipated, with individual substrates being impossible to assay due to interference or lack of solubility. Below in Table 11 and Table 12 are the full data acquired from the substrate screens with ADH D.

Substrate	Specific Activity ($U\ mg^{-1}$)
Butanal	10.4
Decanal	3.7
2-Butanone	-
2-Decanone	-
3-Pentanone	-
Isobutanal	8.5
Cyclohexanone	-
Furfural	6.8
Acetophenone	-
Ethyl acetoacetate	0.5
Benzaldehyde	6.9
Pyruvate	-
Ethyl pyruvate	3.0
Ethyl levulinate	-
Levulinate	-
α-Tetralone	-
2,3-Pentanedione	5.6
Ethyl 2-oxo-4-phenylbutyrate	6.8
4-Phenyl-2-butanone	-
1-Phenyl-1,2-propanedione	11.0
2,2,2-Trifluoroacetophenone	-
Ethyl 4-chloroacetoacetate	2.6
Ethyl 4,4,4-trifluoroacetoacetate	-
2',3',4',5',6'-Pentafluoroacetophenone	-

Table 11 - Primary substrate screen of ADH D, using NADPH as cofactor.

Substrate	Specific Activity ($U\ mg^{-1}$)
Methanal	0.9
2,3-Butanedione	12.5
2,5-Hexanedione	-
2,3-Hexanedione	11.4
2,3-Heptanedione	11.2
3,4-Hexanedione	2.9
Ethyl acetoacetate	0.6
Methyl acetoacetate	0.4
Ethyl 4-chloroacetoacetate*	4.9
Ethyl 2-chloroacetoacetate	9.4
Methyl 4-chloroacetoacetate	3.4
Ethyl 4,4,4-trifluoroacetoacetate*	0.2
Glyoxal	12.3
3-Methylbutanal	-
5-Norbornene-2-carboxaldehyde	12.9
4-Chlorobenzaldehyde	19.4
4-Nitrobenzaldehyde	n.d.
2,4-Dimethyl-3-pentanone	-
Benzophenone	-
Benzoin	Activity
Benzil	-
Acetoin	-

Table 12 - Secondary substrate screen of ADH D. *Repeat from primary screen, benzoin interfered with assay allowing qualitative analysis of activity only. 4-nitrobenzaldehyde interfered with assay.

From the screening it was determined that ADH D was a broad range alcohol dehydrogenase, accepting both aldehydes and certain ketones. Of particular note was the inability to accept basic monoketones, which was highlighted as a potential benefit in the case of cofactor recycling, allowing for sacrificial substrates such as acetone to be utilised in a two enzyme, two substrate cofactor recycling system.

Diketones and chlorinated substrates were accepted well, but fluorinated substrates were poorly tolerated, if accepted at all. Substrate size was not a concern either, with large and bulky substrates such as 5-norbornene-2-carboxaldehyde, 1-phenyl-1,2-propanedione and benzoin being accepted along with glyoxal and even methanal. Interestingly, diketones had to be adjacent; diketones with the carbonyl groups spread across the molecule were not accepted; for example 2,5-hexanedione had no activity with ADH D whereas 2,3-hexanedione did.

Activity appears to be higher with chlorinated substrates, or those with additional localised electronegativity, for example in the case of ketoesters; this is characteristic of alcohol dehydrogenases.

As is typical for an enzyme with a wide substrate scope, specific activities were relatively low. Potentially possible that substrate inhibition was a root cause, but unlikely at 10mM concentrations; in either case, this is something that was examined in an investigation of the enzyme's kinetics.

3.1.1. KINETICS

Several substrates were chosen from the screening on the basis of higher activities amongst the substrates tested. With the assumption that an enzyme would have comparatively high activity with a native substrate or a well-accepted substrate at a concentration of 10mM, it was noted that recorded activities would decrease when considering the kinetics. Therefore, previously obtained rates were considered to be at the upper end of the scale, necessitating utilising substrates with rates unlikely to be reduced to background noise during the course of the investigation.

Due to the low solubility of several substrates in buffer, stock solutions of these substrates were made up in a relevant solvent that dissolved into buffer. Solvents were tested on a substrate by substrate basis, but typically were DMSO, acetonitrile or ethyl acetate, and these solvents comprised a final amount of 5vol% or below, depending on previously determined solvent sensitivity in a preliminary solvent screen. Stock solutions were generally in the range of 100mM to 1M concentrations irrespective of the substrate and solvent. Final concentrations of substrate varied, but the aim was to have ten separate concentrations tested as a minimum. Interim results were graphically displayed, and the experiment ended with a given substrate when either it was apparent that near V_{max} had been obtained when working in increasing concentrations, or where a sufficient curve had been recorded, when working in decreasing concentrations.

Normally substrate concentrations were chosen starting with 10 or 20mM as a maximum concentration, working downwards to complete an entire Michaelis-Menten curve and define the Michaelis constant, but buffer soluble substrates frequently were studied at higher concentrations. On occasion though, these concentrations were insufficient for substrates in solvent and therefore were pushed higher.

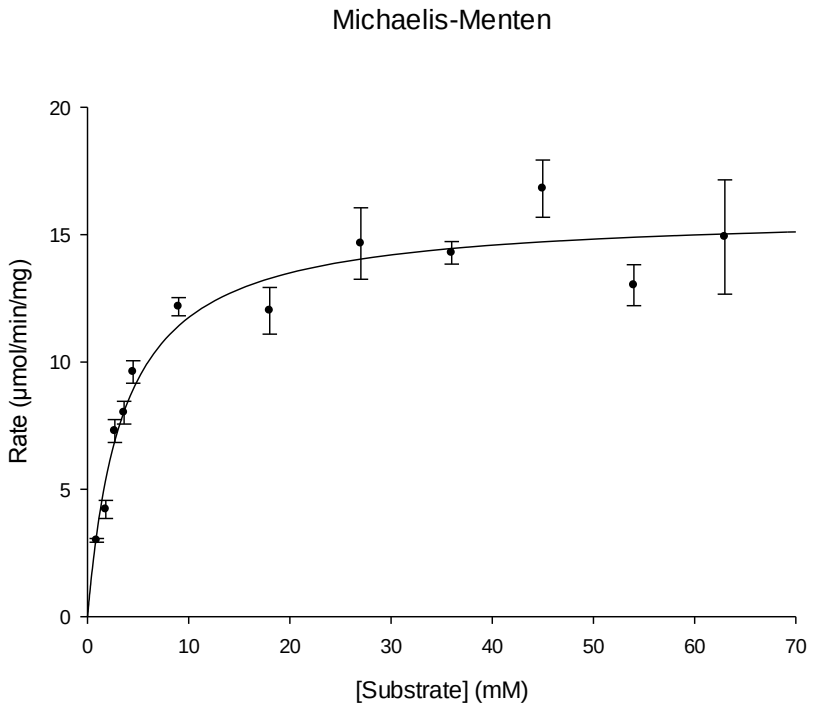


Figure 67 – Graph of rate of reaction vs [S] using ADH D and butanal. Data fitted to a Michaelis-Menten curve equation using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

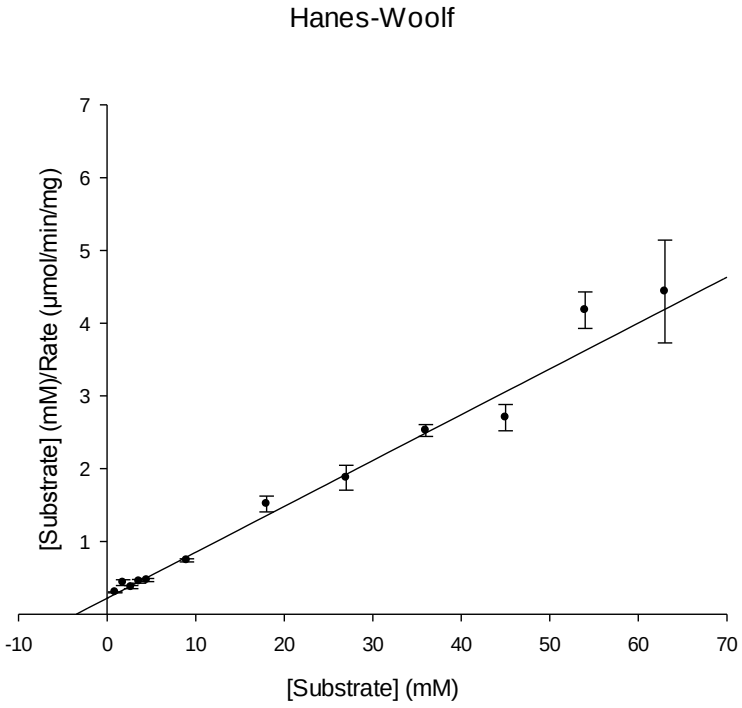


Figure 68 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and butanal. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

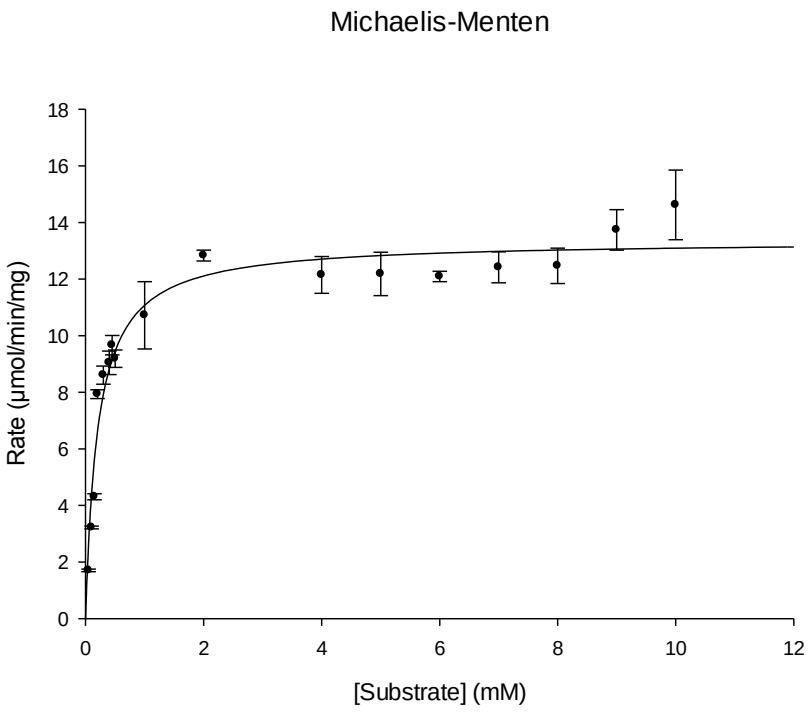


Figure 69 - Graph of rate of reaction vs [S] using ADH D and 1-phenyl-1,2-propanedione in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Figure 70 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 1-phenyl-1,2-propanedione in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

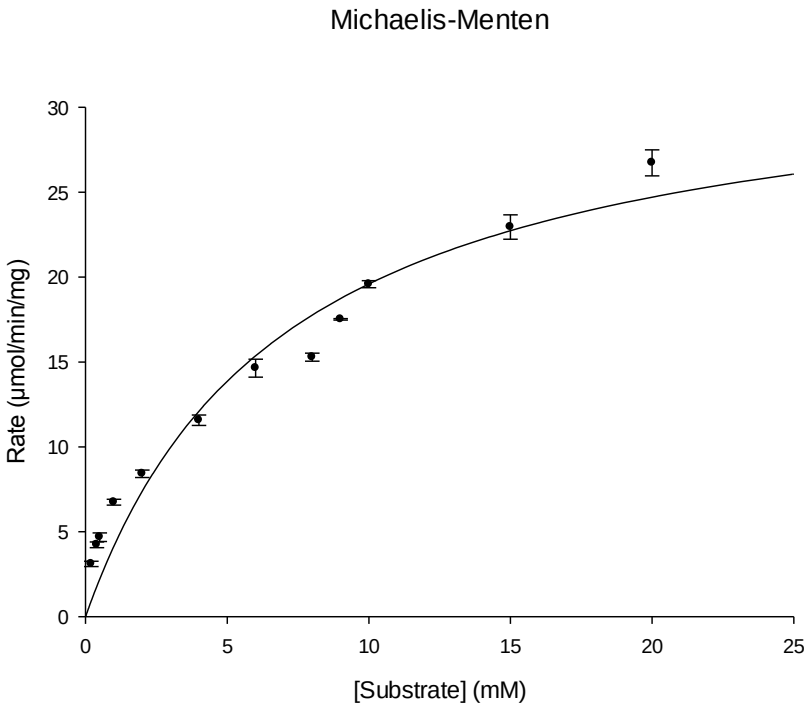


Figure 71 - Graph of rate of reaction vs [S] using ADH D and 4-chlorobenzaldehyde in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

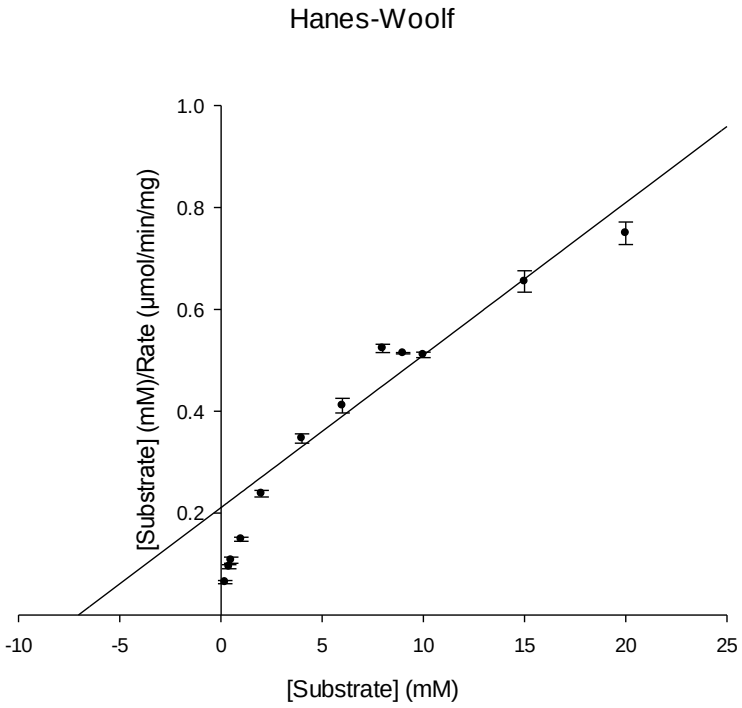


Figure 72 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 4-chlorobenzaldehyde in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

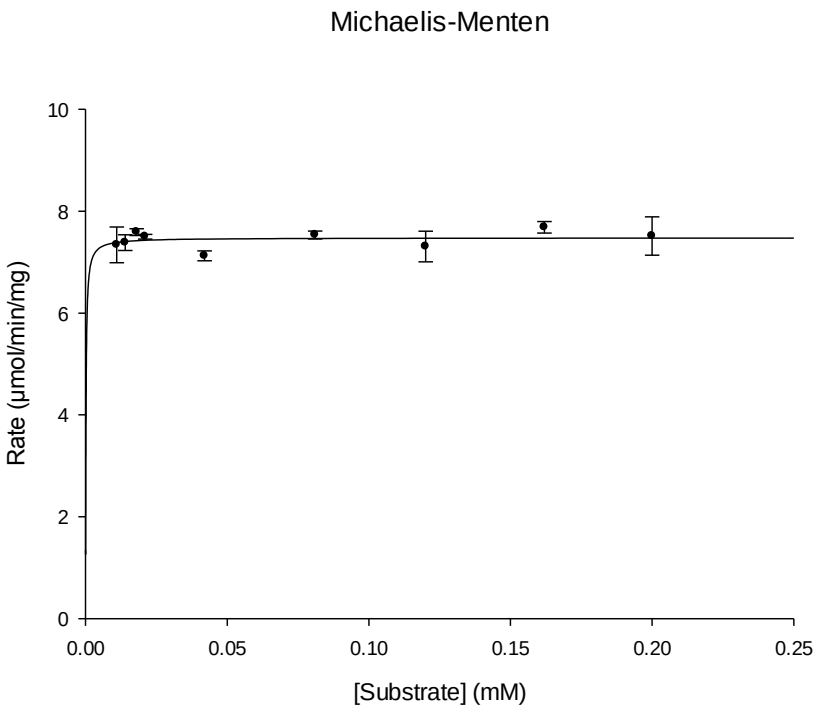


Figure 73 - Graph of rate of reaction vs [S] using ADH D and NADPH. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM butanal, 50 °C, 1ml assay volume.

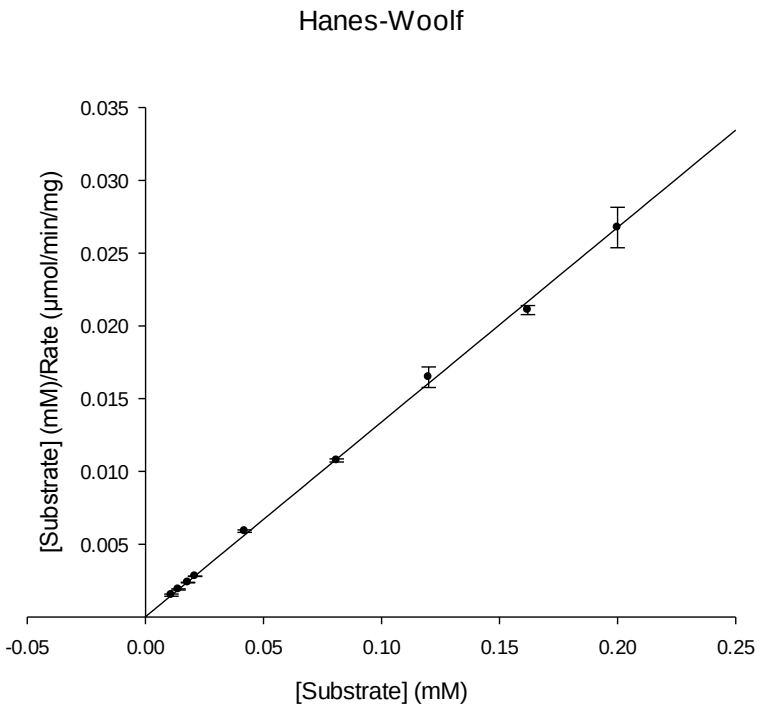


Figure 74 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and NADPH. Conditions: 50mM sodium phosphate pH 7.0, 10mM butanal, 50 °C, 1ml assay volume.

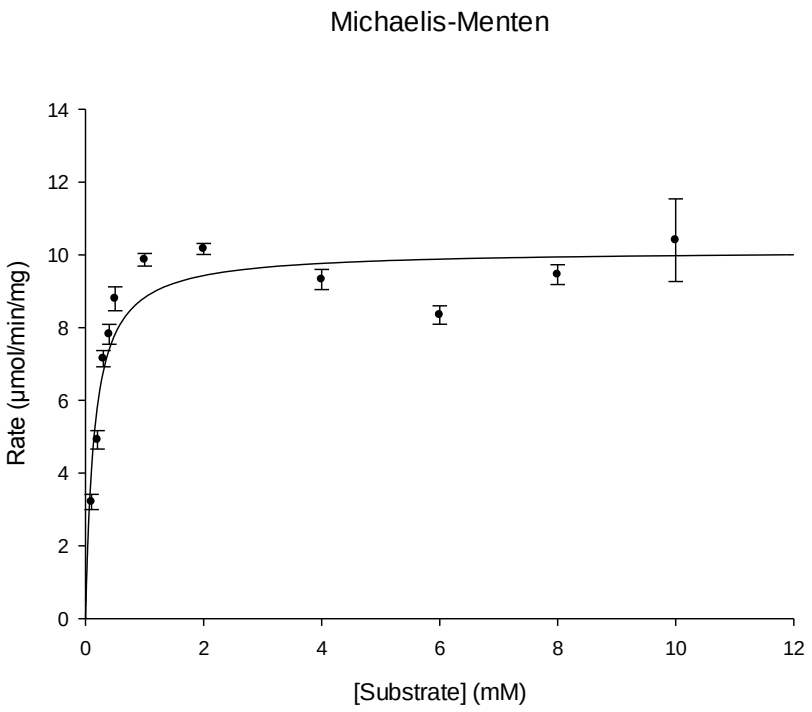


Figure 75 - Graph of rate of reaction vs [S] using ADH D and 5-norbornene-2-carboxaldehyde in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Figure 76 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and 5-norbornene-2-carboxaldehyde in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

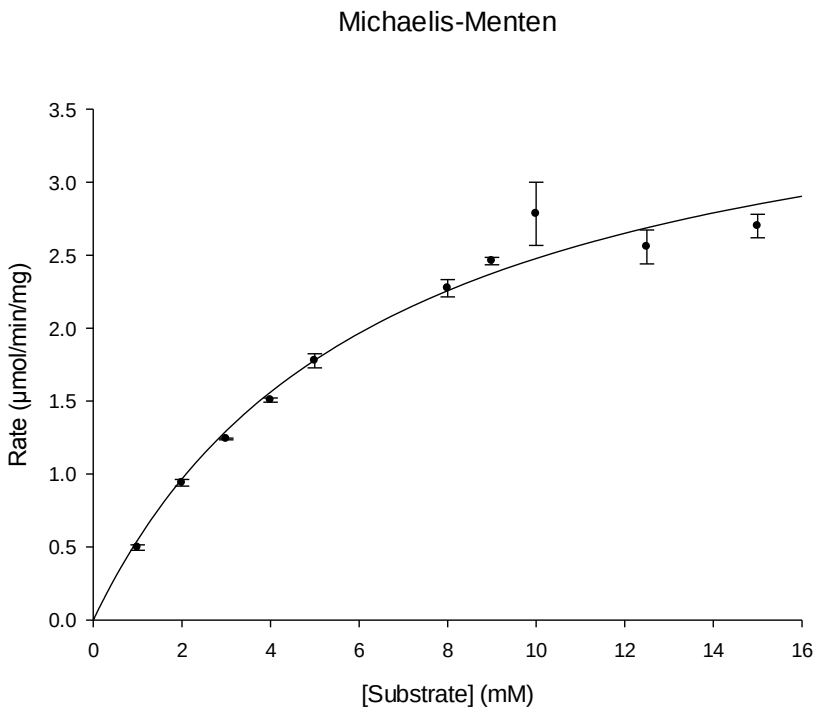


Figure 77 - Graph of rate of reaction vs [S] using ADH D and ethyl 4-chloroacetoacetate in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

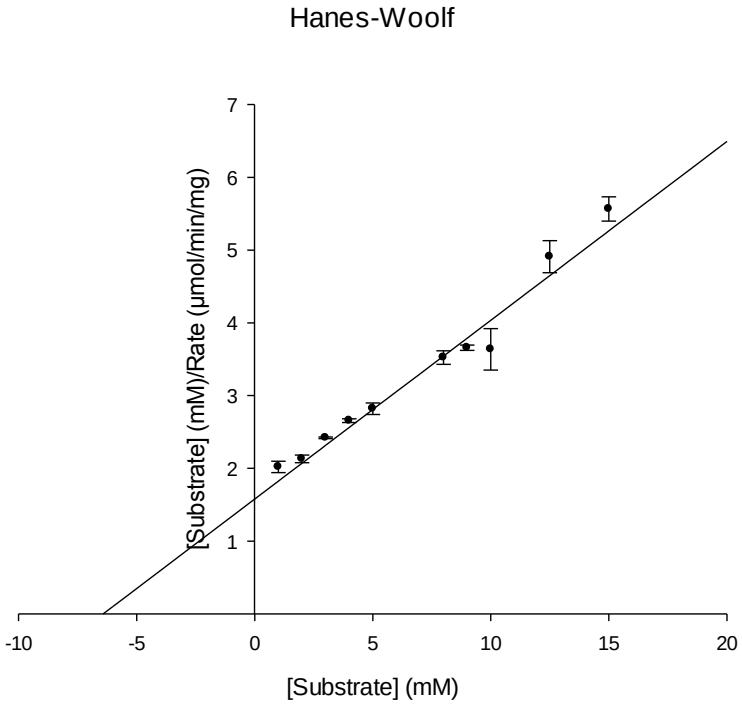


Figure 78 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and ethyl 4-chloroacetoacetate in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

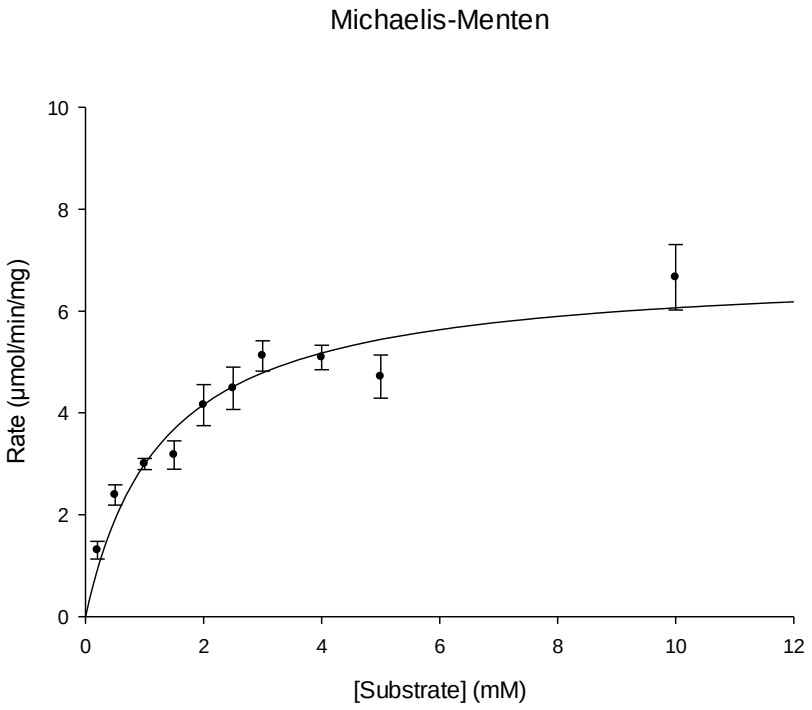


Figure 79 - Graph of rate of reaction vs [S] using ADH D and ethyl 2-oxo-4-phenylbutyrate in ethyl acetate. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Figure 80 - Hanes-Woolf plot ([S]/v vs [S]) for ADH D and ethyl 2-oxo-4-phenylbutyrate in ethyl acetate. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
Butanal	15.86(±0.57)	3.50(±0.54)	600	170
1-Phenyl-1,2-propanedione	13.36(±0.27)	0.21(±0.02)	510	2400
4-Chlorobenzaldehyde	33.41(±2.33)	7.04(±1.15)	1300	180
NADPH	7.47(±0.11)	1x10^{-4}(±3x10^{-4})	290	2800000
5-Norbornene-2-carboxaldehyde	10.12(±0.30)	0.14(±0.02)	390	2700
Ethyl 4-chloroacetoacetate	4.07(±0.24)	6.42(±0.85)	160	24
Ethyl-2-oxo-4-phenylbutyrate	5.97(±0.38)	0.88(±0.19)	230	260

Table 13 - Summary of kinetic data with ADH D.

All substrates chosen for ADH D have some industrial relevance, with the exception of butanal, which is an example substrate. Whilst this ADH will only accept NADPH rather than the less costly NADH as a cofactor, it was not possible to determine accurate kinetic constants due to the incredibly low K_M; due to the assay measuring decrease in absorbance of the cofactor there is a natural limit on the level of cofactor that can be realistically examined. Although speculative, the estimated specificity constant for the cofactor does not exceed the diffusion limit and therefore should not be immediately discounted due to its apparent size. In particular, as the affinity for cofactors in general is much higher than substrates and given that enzymes using NADPH generally have lower Michaelis constants than those using NADH, a specificity constant of 2,800,000 mM^{-1} min^{-1} remains plausible (Bar-Even et al., 2011).

Interestingly, although it appears that this enzyme has a greater affinity for larger substrates, this is not the case. For example, the highest affinity noted was for 5-norbornene-2-carboxaldehyde, yet ethyl 2-oxo-4-phenylbutyrate is a larger molecule and ADH D has a lower affinity for this substrate; indeed the specificity constant is a full order of magnitude below. Lastly, given that butanal has a similar specificity constant to ethyl 2-oxo-4-phenylbutyrate and is the smallest substrate tested at this stage, there does not appear to be any particular bias towards larger substrates in this regard.

As all these substrates aside from butanal are either aromatic or are halogenated, or both, there is no obvious pattern with regards to electronegativity near the target carbonyl group, with both phenyl and chlorinated substrates not noticeably affecting the kinetic parameters. Frequently ADHs are more active with substituted substrates, particularly halogenated substrates, and ADH D appears to be following this trend. Helpfully, this is beneficial as traditional chemical transformations typically are challenged by additional substituents, either not working as required or requiring additional method steps for adequate protection to ensure the desired product.

3.1.2. STABILITY ANALYSIS AND TERTIARY SCREENING

In the case of thermostability, aliquots of ADH D were tested for their resilience at temperatures above the standard growing temperature of *G. thermoglucosidasius*, which is 60°C. No activity was recorded at 81°C or 90°C; however, a steady reduction in activity was noted at 70°C. In fact, an apparent two step reduction was noted, with a faster reduction in

the first hour, which then stabilises to a near full retention of activity thereafter. This could indicate a multiple step denaturation process whereby the first stage results in a partially active enzyme structure and the second stage is a full denaturation where all activity is lost.

Importantly, the first step must be irreversible as samples were not tested immediately after incubation and were all kept on ice until testing. As both preliminary work with hexanal and a repeat with furfural record this "two step", with both indicating a faster reduction in activity over the first hour; this raises confidence that this is not an artefact of one particular experimental run or substrate.

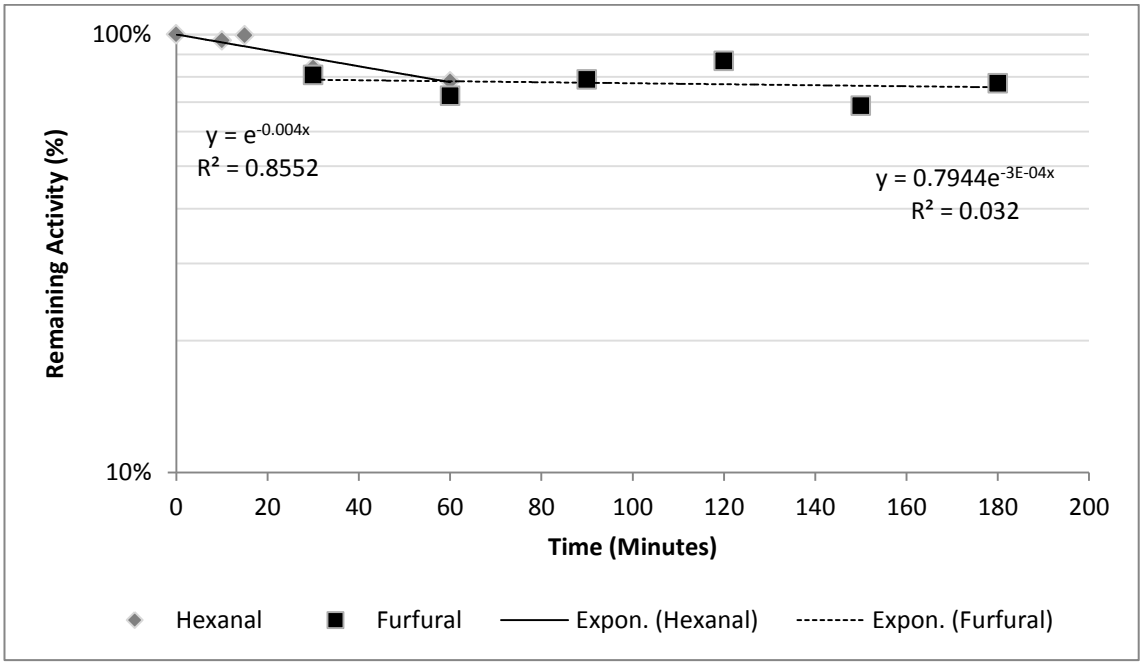


Figure 81 - Thermostability data acquired storing enzyme samples at 70°C in PCR machine. No activity noted after 5 minutes at 81°C or 90°C; hexanal and furfural used as substrates, NADPH used as cofactor. Zero point removed from furfural data for trend line purposes.

Another potential explanation for the above data is that the initial reduction is caused by a denaturation of a second enzyme, a mesophilic host ADH. Whilst this is theoretically possible, the ADH was electrophoretically pure, as shown in Figure 66, therefore this is unlikely to be the case.

Assuming the second rate is the longer term "stable" reduction in rate of activity, then the estimate of the half-life becomes:

$$t_{1/2} = \frac{ln(2)}{0.0003}$$

$$t_{1/2} = 2{,}310\ mins$$

With an estimated half-life of 2310 minutes, or 38.5 hours, at 70°C it is clear that ADH D is still operational at the upper limits of *G. thermoglucosidasius* growing temperature. Similarly, with no activity recorded at 81°C, which is above the temperature at which the organism can live, there appears to be little discrepancy between the enzyme and species.

The poor correlation between the trend line calculated in Figure 81 and the raw data for thermostability of ADH D is concerning but likely to be caused by the relatively low degradation rate combined with a high rate of noise. Thus additional work with lengthier timescales, perhaps over days, would reveal whether the estimate is overstating or even understating the half-life of this enzyme.

Because of ADH D's interesting substrate scope and potential for industrial application, a full tertiary screen was conducted. This screen comprised an extensive list of potential denaturants, buffers, solvents and ions to provide some indication of ADH D's operational capacity. In essence this is a feasibility study into the ability of ADH D to maintain activity in a range of situations as well as a partial optimisation with respect to buffer/pH and additive ions.

Assay Condition	Relative Activity (%)
pH 5.0 Sodium Acetate	83
pH 5.0 Succinate	94
pH 6.0 Citrate	109
pH 6.0 Succinate	104
pH 6.5 MES	82
pH 6.5 Citrate	112
pH 7.0 Sodium Phosphate	100
pH 7.0 MOPS	87
pH 7.5 Sodium Phosphate	85
pH 7.5 Imidazole	95
pH 8.0 Sodium Phosphate	67
pH 8.0 HEPES	57
pH 8.5 Borate	56
pH 8.5 Bicine	49
pH 9.0 Borate	40
pH 9.0 Bicine	35
Unbuffered	108

Table 14 - Tertiary screen with ADH D (buffers) - all tested with 10mM butanal as substrate, averages of duplicate assays.

As expected, lower pH was a benefit to the studied reaction whereas alkaline pH values caused a relatively steep drop in recorded activity. That said, the enzyme was relatively insensitive to the buffers used, with only assays run in MES having a notable reduction in activity with a 30% differential between the MES assays and those run in citrate. Unsurprisingly, the highest activity recorded was at pH 6.5, using citrate as buffer. As a relatively benign buffer and its wide buffering range, this is very useful. It should also be stated that it is valuable to know that chelation is not a major issue for ADH D.

Additive	Dose (vol%)	Relative Activity (%)
10% SDS	1	-
500mM EDTA (pH8.0, water)	0.2	96
	2.0	88
Triton X-100	1	93
	10	92
PEG 600	1	99
	10	75
DMSO	1	85
	5	72
	10	58
	20	33
	50	-
Ethyl Acetate*	1	102
	5	114
	10	105
	20	105†
Acetonitrile	1	99
	5	115
	10	120
	20	95
	50	27
Cyclopentyl Methyl Ether*	1	102
	5	100
	10	96
	20	102†

Table 15 - Tertiary screen with ADH D (additives and solvents) - All tested using 10mM furfural as solvent and pH 7.0 sodium phosphate. *Ethyl acetate and CPME form bilayers, impacting upon assays. For example, ethyl acetate above 8% forms a bilayer, indicating solvent saturation, potentially affecting observed rates and at high levels distorting assays. All values are averages of duplicate assays, except those marked †, which are single assays and have solvent effects on the assay.

Several generic additives were tested to examine their effects upon ADH D, as well as several solvents. In the case of the additives, it appears that a low quantity of a benign substance such as PEG 600 is well tolerated; however, greater volumes may prove problematic, with a 25% reduction in activity upon the addition of 10vol%. Predictably, the addition of a more aggressive additive such as SDS resulted in the loss of all activity. A small loss of activity (~7-8%) was noted with the addition of another surfactant, Triton X-100, indicating that at least some surfactants may be well tolerated by ADH D. Given that SDS is an anionic surfactant and Triton is a non-ionic surfactant, it seems likely that this is the cause of the difference in tolerance.

The high acceptance of EDTA provides further evidence that ADH D is resistant to the chelation of metals, either indicating that it does not rely upon the use of metal ions such as zinc or iron for activity, or that any metal ions relied upon are very tightly bound to the enzyme and cannot be easily abstracted by chelating agents.

Ion	Dose (vol%)	Relative Activity (%)
100mM $FeCl_2$	1	91
	10	75
100mM $ZnCl_2$	1	85
	10	88
100mM $MnCl_2$	1	92
	10	87
100mM $MgCl_2$	1	89
	10	93
100mM $NiCl_2$	1	89
	10	90

Table 16 - Tertiary screen with ADH D (ions) - All tested using 10mM furfural and pH 6.5 citrate buffer. Necessary to alter buffer composition due to reaction of ions with sodium phosphate buffer. Note high absorbance of ions at 10mM Concentration.

In order to test the addition of ions as additives to the assay, a change of buffer was needed as sodium phosphate reacts with the chlorides forming [ion] phosphate and an opaque assay due to the insolubility of those phosphates. From the previous work citric acid was chosen as the optimal buffer studied to carry out this work. With the benefit of hindsight, this was an unsuitable buffer to choose due to the aforementioned chelation ability of citric acid. However, it is unlikely that all of the relevant ion is chelated. Therefore, the concentration of ion expected is likely to be an overestimate. As the original concentrations of added ions are rather high, this is less of an issue than it could be, although the quantification of the additives is now impossible.

Buffer suitability aside, it appears that there is little effect of additional ions, with the exception of the high dose of iron. Given that previous additives such as EDTA had little effect it was expected that metal ions were not required for activity, and therefore the addition of more should have little effect. This is precisely what is seen here.

Another important point is that there is likely to be traces of nickel present as part of the purification using a nickel based column. Although only a very small concentration could be present, this could be sufficient to cause issues when considering an equally low concentration of enzyme, which is only present at nanomolar or micromolar concentration when purified. Due to the relative concentrations of the citric acid and added ions though, this source of contamination can be entirely discounted. The quantity of nickel present is orders of magnitude lower than both the citric acid and the added ions, which are approximately the same; the lower concentration is five times less than the citric acid, and the higher roughly double.

3.1.3. OPTIMAL TEMPERATURE AND MELTING POINT

Another area of study is the determination of the optimal temperature and melting point of ADH D. Whereas previously the residual activity of the enzyme was calculated after a specific length of time incubated at high temperatures, cooling down and then assaying at a fixed temperature, the optimal temperature is investigated by running identical assays at a range of temperatures to record the specific activity obtained.

As there is a general increase in activity as the temperature is increased, roughly doubling every ten degrees increase, and there is a known point at which the enzyme activity is no

longer retained, then there will be a point whereby the degradation of the enzyme outweighs the steady increase in activity, and an inflection point will be found, which is termed the optimal temperature. Whilst it is unlikely that any process would run close to the optimal temperature, despite the name, it is a convenient piece of information to know as the absolute fastest rate can be determined. Further, based upon the data from the temperature screen, it will be possible to estimate the rate of reaction at any given temperature. Combined with stability information at the same temperature it would be possible to use the information gathered to optimise a process for a given substrate at a required temperature.

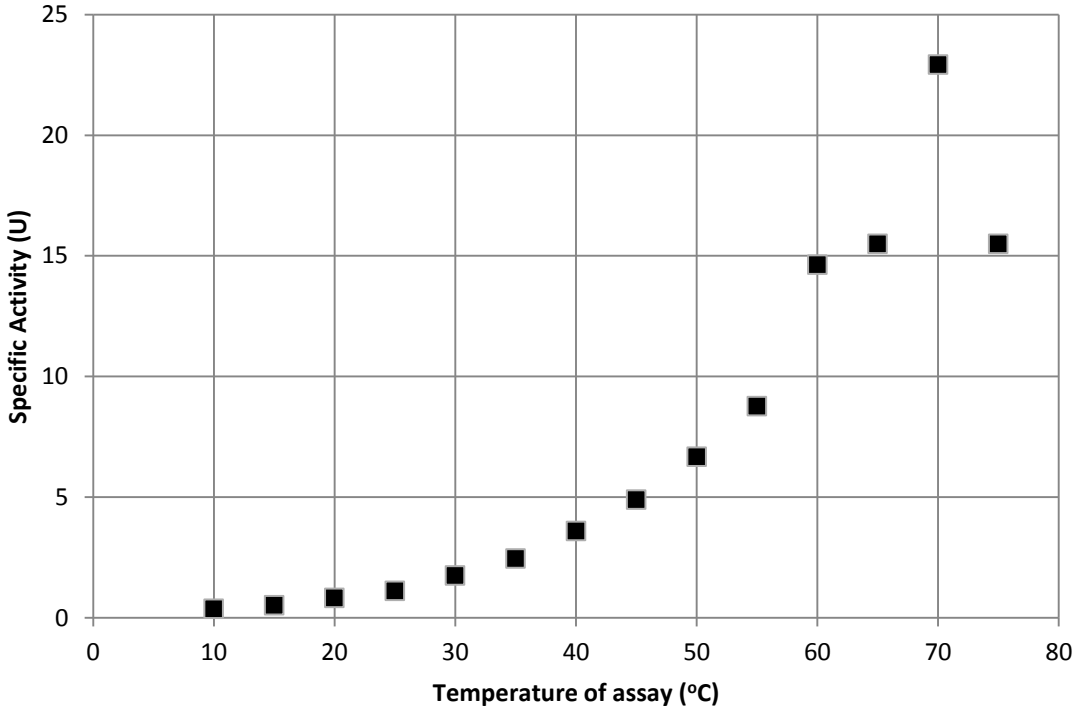


Figure 82 - Optimal temperature study with ADH D, using butanal as substrate and NADPH as cofactor.

The figure above indicates the likely optimal temperature to be 70°C, which is appropriate in view of the previous thermostability analysis and the growth range of *G. thermoglucosidasius*. Whilst the boiling point of butanal, the substrate used, is 74°C, the length of time of a standard assay is unlikely to affect proceedings very much; the study was using 10mM butanal, approximately $3K_M$. Between the high solubility of the substrate and the brevity of the assays including the preincubation phase and the covering of the cuvette during the assay, evaporation can be neglected.

A further analysis was conducted to evaluate the thermal stability of enzymes. Called a thermal shift analysis, the technique exploits a dye that is able to bind to hydrophobic residues. Upon binding it is strongly fluorescent and is then strongly quenched by water. Thus, by monitoring excitation at 470nm and emission at 550nm, it is possible to identify when proteins are denaturing by considering the derivative of the fluorescence. When the inflection point of the derivative is reached, i.e. when the increase in dye binding has reached and passed its maximum, this is the melting point. This depends entirely on the easily detectable increase in fluorescence when the dye binds to the protein.

The results from this revealed two inflection points, at 76.7°C and 81.2°C, indicative of a two stage denaturation process. The thermostability investigation determined that there was no activity after even five minutes incubated at 81°C, but substantial activity after incubation at 70°C, without intermediary points. Therefore further work would have to be done to identify

the significance of a two-step denaturation system. It is likely that the first step is a reversible stage, with the second being irreversible. As the dye method would probably preclude attempting to partially unfold the enzyme by heating part way and cooling, the best method of investigating this is likely to be dependent on the assumption that activity will correlate with structure. If heating the enzyme to 77^{o}C briefly before cooling allows for activity to be retained, and heating to 81^{o}C does not, then it is likely that the first stage is irreversible. If activity is not retained, then this will not indicate anything as it is possible for activity to be lost without an irreversible conformational change.

It is crucial to note that this method only reveals the exposure of hydrophobic regions of the enzyme which can bind to the dye, and the effectiveness of binding. Whilst this is generally sufficient because the exposure of hydrophobic regions of a protein, normally deeply buried, is correlated with and taken to imply denaturation, some ambiguity can remain. For instance, initial exposure of hydrophobic areas could actually indicate a breakdown of polymerisation, i.e. a previously dimeric or multimeric enzyme initially loses coherency and breaks down to a monomeric form. This enzyme may or may not be active as a monomer, perhaps even more active, but in all cases would not represent full denaturation of the enzyme, merely loss of quaternary structure. The only way to definitively determine irreversible denaturation is via other structural work such as circular dichroism or differential scanning calorimetry.

3.1.4. CHEMICAL ANALYSIS

In the first instance, several assays were analysed by GC/MS in an end point analysis to provide proof of the products and a quantitative measure of conversion. There is no difference between the assays studied by this method to those previously run in a cuvette using a spectrophotometer with the exception of the cofactor regeneration system used. A two enzyme, two substrate system was utilised via the addition of a thermostable glucose dehydrogenase and glucose to the assay. The data are summarised in Table 17.

	Conversion	
Substrate	**1 Hour**	**16 Hours**
5-Norbornene-2-carboxaldehyde*	95%	74%
Butanal**	31%	17%
Furfural	51%	59%
Methylbenzoylformate	-	-
Ethyl-2-oxo-4-phenylbutyrate	49%	65%
4-Phenyl-2-butanone	0%	0%
Decanal	3%	Activity
Benzaldehyde	1%	Activity
4-Chlorobenzaldehyde	31%	63%
1-Phenyl-1,2-propanedione†	24%	29%/9%
2',3',4',5',6'-Pentafluoroacetophenone	2%	5%
Ethyl 4-chloroacetoacetate	N/A	-
Ethyl acetoacetate	N/A	4%
Acetophenone	N/A	0%
2,3-Pentanedione	N/A	10%/38%
2,3-Butanedione	N/A	56%

Table 17 - Summary of GC/MS analysed assays. All assays conducted for one hour or at 16 hours at 50°C. N/A indicates not tested, "-" indicates unsuitable substrate for GC/MS analysis. *Used 20µl instead of 10µl of enzyme in 1 hour assay. **Suspected evaporation of substrate. †Two products detected, individual conversions provided after 16 hours.

As the substrates tested by GC/MS were chosen on the basis of activity among all ADHs tested, there are some that were expected to have no activity, for example acetophenone. Methylbenzoylformate is a substrate of industrial relevance that was unsuitable for testing by spectrophotometer due to absorbance at the same wavelength as NAD(P)H, but was analysed by GC/MS as an alternative method of testing; however this also proved unsuitable.

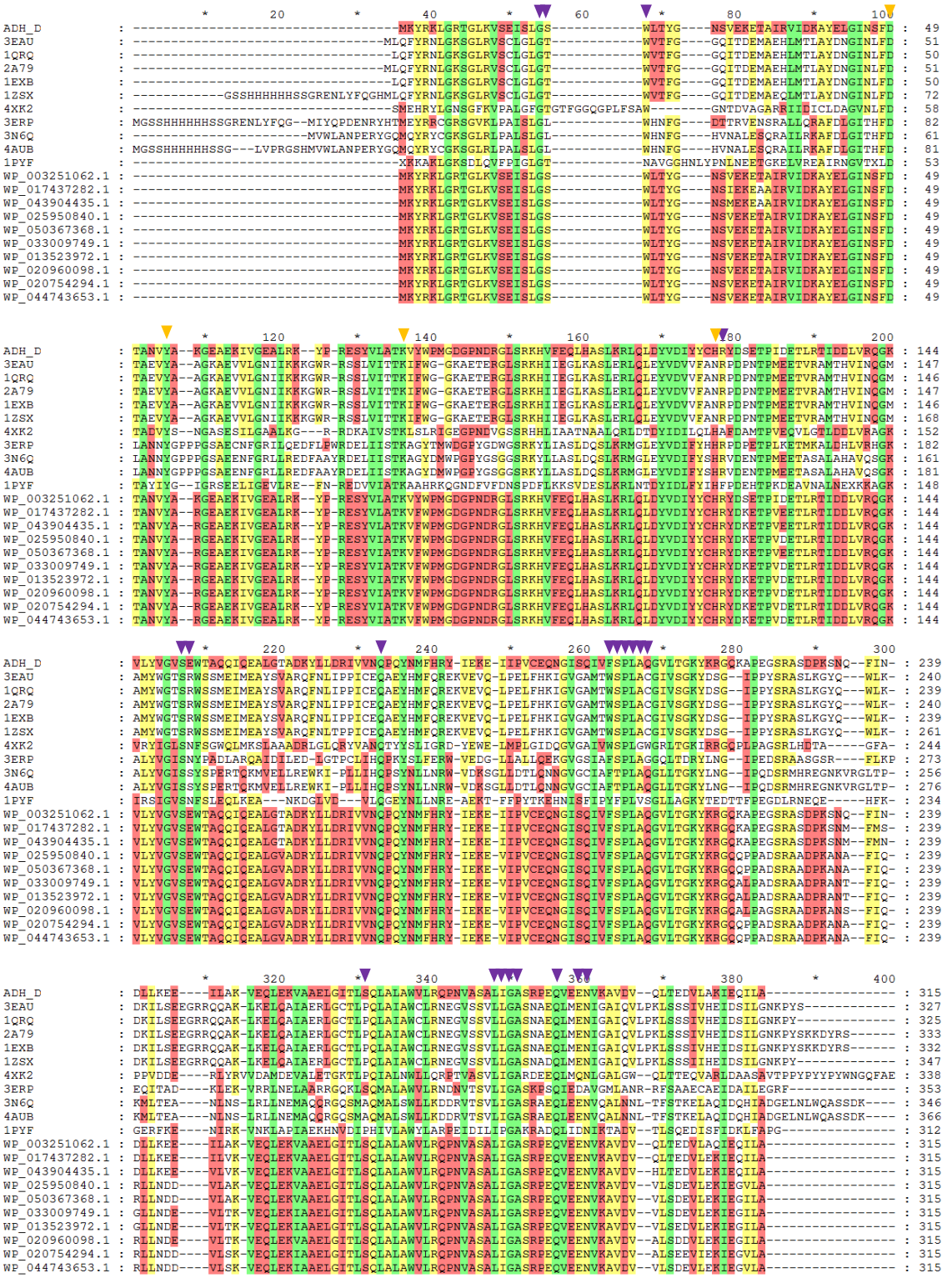
Overall there is only one surprise in terms of activity, and that is 2',3',4',5',6'-pentafluoroacetophenone, which was believed to have no activity after assaying, but product was detected by GC/MS. In general those substrates with which ADH D had the highest initial rate had higher conversions after a longer period of time, although none demonstrated quantitative conversion. It is likely that additional conversion would be noted by using higher concentrations of enzyme, something apparent from the 5-norbornene-2-carboxaldehyde entry where double the enzyme was inadvertently used for the one hour assay, resulting in a greater conversion than the 16 hour assay. It is not immediately clear whether the enzyme or another aspect of the assay degrades over the course of the 16 hour period, but it is likely that cofactor degradation is the root cause; at elevated temperature the half of life of NADPH can range from under three hours to over 28 hours in phosphate buffer, depending on the ionic strength (Wu et al., 1986). Those estimates were at 41°C, so they are liable to be overestimates. It would also seem that NADH is almost invariably more stable than NADPH, another reason for its preference in industrial settings.

When diketones were tested as substrates, two products were detected, both expected products if only one ketone were converted to an alcohol, and no diols were recorded. For example, in the case of 2,3-pentanedione, both 2-hydroxy-3-pentanone and 3-hydroxy-2-pentanone were detected, but no 2,3-pentanediol. A particular problem with the variety of potential products from using a diketone as substrate is that the usual external standard method was ineffective, i.e. using a known concentration of the product to produce standard curves with which to measure the quantity of the unknown conversions. In this case, an assumption was made that both forms of the singly transformed product were equivalently detected by the GC/MS in order to convert the detected signal into a useful form. This assumption produced good mass balances, certainly as good as those substrates with relevant standard curves, if not better.

The detection of both singly converted products implies that ADH D is promiscuous with respect to accepting either ketone as a substrate, but unable to utilise the second ketone on the same molecule. This is particularly surprising in the example of 2,3-butanone, as the only product detected was acetoin. Given the symmetrical nature of the substrate, it is surprising that a hydroxyketone is not an acceptable substrate.

If the relative proportions of the products gleaned from diketone substrates represent final conversions, ADH D has a bias of 3:1 in favour of 1-hydroxy-1-phenyl-2-propanone compared to the alternative, and a bias of 4:1 in favour of 2-hydroxy-3-pentanone. Whilst a less than exemplary result, it is at least better than a complete lack of bias and perhaps could be improved at a later stage.

3.1.5. SEQUENCE ALIGNMENT



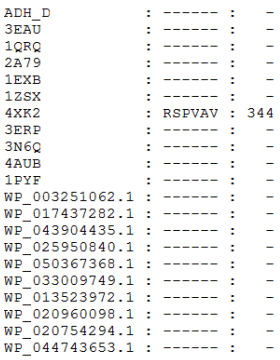


Figure 83 - Multiple Sequence Alignment of ADH D with most similar protein sequences associated with crystal structures and obtained from the PDB and most similar protein sequences obtained from the NCBI RefSeq Database using NCBI Seqr Sequence Search. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to crystal structures from the PDB or proteins from the NCBI RefSeq Database. Purple triangles indicate amino acids involved in the active site and orange triangles indicate the catalytic tetrad as indicated by the NCBI CDD. All these sites provided via the NCBI Conserved Domain Database and obtained through the Conserved Domain Search. All protein sequences annotated as Voltage-Gated potassium channel beta subunit from the genera *Geobacillus* and *Anoxybacillus*; the crystal structures are as follows: 3EAU, 1QRQ, 2A79, 1EXB are all voltage-gated potassium channel beta subunit from *Rattus norvegius*, 1ZSX is a voltage-gated potassium channel beta subunit from *Homo sapiens*, 4XK2 is an AKR from *Polaromonas sp. JS666*, 3ERP is a putative oxidoreductase from *Salmonella typhimurium*, 3N6Q is the crystal structure of YghZ, a methylglyoxal reductase from *E. coli*, 4AUB is an AKR (AKR14A1) from *E. coli*, 1PYF is an AKR (AKR11A) from *B. subtilis*. Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997).

The four closest crystal structures appear to be the same protein crystallised at different times and, thus shall be considered a single entry for the purposes of discussion. Work relating to 1ZSX, 4XK2 and 3ERP remains unpublished and therefore cannot be examined beyond visual inspection of the MSA.

3EAU is a member of the Shaker family voltage-dependent potassium channel, which are critical for allowing the flow of potassium ions out of cells, modulating electrical activity. In humans, loss of function mutations have been associated with serious conditions including atrial fibrillation. In addition, two of three mammalian beta subunits of the channel complex are also tetrameric NADPH-dependent AKRs of unknown function (Long et al., 2005; Pan et al., 2008). Work done with 2A79 has elucidated the probable *in vivo* function of the AKR as modulating the inactivation of the potassium channel and identified activity with 4-carboxybenzaldehyde and 4-cyanobenzaldehyde. Further, the catalytic tetrad is identified as D85, Y90, K118 and N158, corresponding to D100, Y105, K136 and N176 in Figure 86 (Weng et al., 2006). This matches with the residues referred to by the NCBI database, but in the case of ADH D, it is H176, not N176, which is conserved in all AKRs in the MSA aside from voltage-dependent potassium channel beta subunits. This is consistent with some AKR families which differ in their catalytic residues at this position, with the histidine being crucial to substrate specificity. This position remains conserved as either a histidine or a asparagine in all AKR families (Weng et al., 2006).

Two differences have also been noted between traditional AKRs and those found as part of potassium channels. The first is that the cofactor is much more exposed to free solvent in AKRs other than the potassium channel and the second is that AKRs typically have a more enclosed active site (Gulbis et al., 1999). This could be the reason behind the poor activity noted with the substrates noted above, a second reason could be that the AKR in this case is not the primary purpose of the protein complex, i.e. the purpose of the enzyme is not to

produce aldehydes, ketones or alcohols, but rather to act as a method of oxidising the NADPH as necessary (Gulbis et al., 1999). One theory is that the redox state of the cell is the trigger with 4-oxononenal, a product of oxidative stress, also having activity (Weng et al., 2006).

Given the painfully slow rates of reaction (k_{cat} of 0.073 and 0.082 min^{-1} with 4-carboxybenzaldehyde and 4-cyanobenzaldehyde respectively) and the focus with these proteins on their medical implications instead of their biotechnological usefulness (Weng et al., 2006), it is probably more instructive to consider the remainder of the crystal structures. That said, there are some key residue differences which may be important as they are annotated as forming part of the active site, and thus may be the reason for the poor activity. Using ADH D as reference point these residues are E208R, Q268C and S331P and the residues found in ADH D are conserved in the most similar proteins, albeit their annotation is also voltage-gated potassium channel beta subunits. Given the general pattern of residues that are conserved amongst ADH D, the proteins, and the most similar crystal structures (3EAU, 1QRQ, 2A79, 1EXB and 1ZSX) it appears that the annotation may be correct and there is a difference between voltage-gated potassium channel beta subunits from microbial and mammalian sources. Yet as this particular protein is only found in plants, insects and mammals (Weng et al., 2006), and the gene encoding ADH D is located in a gene cluster annotated to be associated with sugar uptake and transport it seems more likely that potassium and sugar channels have structural similarities and that those microbial proteins annotated as involved in potassium transport are in fact involved with sugar transport. Speculatively this could imply that ADH D's activity is incidental to *G. thermoglucosidasius* and its main purpose is to allow for the transport of sugars and the AKR function is to recycle cofactor as required. Alternatively, the structural similarities could have no implication on function and the *in vivo* function for ADH D is as an aldose reductase.

3N6Q and 4AUB are the apo and holo forms of an octomeric methylglyoxal reductase from *E. coli*, alternatively known as YghZ and AKR14A1. It can catalyse a range of substrates including methylglyoxal and isatin, a diketone, and has been implicated in the metabolism of glycerol (Totir et al., 2012; Lapthorn et al., 2013). This range of substrates is notably similar to the range determined with ADH D.

Lastly 1PYF is an AKR from *B. subtilis*, known as iolS or AKR11A. It was believed to be involved in the breakdown of inositol and other polyols; however, no activity was detected with inositol, glucose, fructose or xylose. Activity was detected with DL-glyceraldehyde, D-erythrose, 9,10-phenanthrenequionone and methylglyoxal but Michaelis constants were in excess of 200mM and k_{cat} 2-3 sec^{-1}, making it a rather poor enzyme (Ehrensberger and Wilson, 2004).

A peculiarity of this enzyme is at position 68 in Figure 83, annotated as an active site residue. In 1PYF it is asparagine, whereas in all other cases it is tryptophan. This results in a drastically different positioning of the cofactor, which is only permitted due to the removal of the bulky tryptophan. Ultimately this results in a reversal of the facing of the cofactor, resulting in the *4-pro-S* hydrogen being positioned towards the presumed substrate binding site rather than the *4-pro-R* hydrogen more typical of AKRs (Ehrensberger and Wilson, 2004).

Another key feature is that the C-terminus of 1PYF is unusually short, with additional residues more frequently being noted, forming a 14 residue C-terminal loop which is believed to interact with the active site, perhaps guiding or otherwise manipulating substrates into the active site. The lack of loop opens the active site, and the few enzymes lacking this loop appear to have poor activity, similar to the potassium channels previously discussed. This is compounded by an active site that remains predominantly hydrophobic even though solvent accessible. One example is that AKRs have a key catalytic step in which a lysine residue

activates a tyrosine residue by depressing its pK_a, and thereby allowing proton donation. In 1PYF a water molecule inserts between the two residues which substantially hinders the activating effect (Ehrensberger and Wilson, 2004). Thus it might be beneficial to determine activity with 1PYF in a hydrophobic environment to see if an increased rate of reaction occurs.

A third feature is that typically AKRs have a "safety belt" which holds the cofactor in place via hydrogen bonds, resulting in conformational change upon binding of the cofactor. This is almost entirely absent in 1PYF (Ehrensberger and Wilson, 2004). Limited activity with NADH is attributed to E217, corresponding to E277. This is a unique residue among the enzymes in the MSA, with R277 being the most popular.

Further residues of interest include positions 204, 205, 206, 208, K214 and K281, corresponding to positions 264, 265, 266, 268, K274, K352. The latter two lysines are not previously annotated as involved with the active site. Importantly 1PYF is unique in the MSA at some of these positions, specifically S264F, Q268V and S352K. The significance of these changes cannot be predicted but taken together are likely to be substantial given that they are associated with the active site and particularly the cofactor. The most obvious speculation is that these residue changes are part of the difference in activity between ADH D and 1PYF.

To conclude, ADH D appears to be the founding member of a new subfamily of AKRs, specifically AKR6D1. As such the most similar crystal structures have low sequence identity, being annotated as other subfamilies of AKRs, notably AKR11 and AKR14. Despite annotation suggesting that ADH D is a voltage-dependent potassium channel beta subunit, a more appropriate role is likely to be in mannitol or other sugar uptake or metabolism. That said, insights from related sequences imply that the C-terminus is important for activity and that position 68 is of interest when determining enantiospecificity. To elucidate whether either insight is correct it would be necessary to obtain a crystal structure of ADH D directly. Should no C-terminal loop be present, it may then be possible to add one based upon the existing sequences.

With reference to *in vivo* activity and function the intriguing possibility is that AKR activity is incidental rather than the primary function. This proposal depends on the assumptions that ADH D has an *in vivo* role of sugar uptake regulation and that said function operates in a similar manner to that of a potassium channel. This may be an area for future investigation.

3.1.6. SUMMARY OF ADH D

ADH D is a classic alcohol dehydrogenase with broad acceptance of aldehydes and ketones. Whilst unable to utilise basic ketones, the enzyme is able to accept branched aldehydes, cyclic and aromatic aldehydes, complex ketones and basic diketones. Chlorinated substrates are well accepted, but fluorinated substrates much less so. With demonstrated activity utilising substrates of a very wide range of molecular mass and composition including several of industrial interest such as 5-norbornene-2-carboxaldehyde and 1-phenyl-1,2-propanedione, ADH D has shown promise.

A potential operating temperature of 70-75^{o}C and wide buffer and pH toleration are benefits of this enzyme, and NADPH and non-specific product formation when using diketones are the challenges of this enzyme.

It is likely this enzyme will be of most assistance when utilised with large substrates with multiple functionalities but only one available ketone group.

3.2. ADH E

The second of two enzymes produced in the initial batch, ADH E was under investigation as it was thought to be to be involved with propanal degradation. No problems were noted with protein expression or purification, with an example SDS-PAGE gel below.

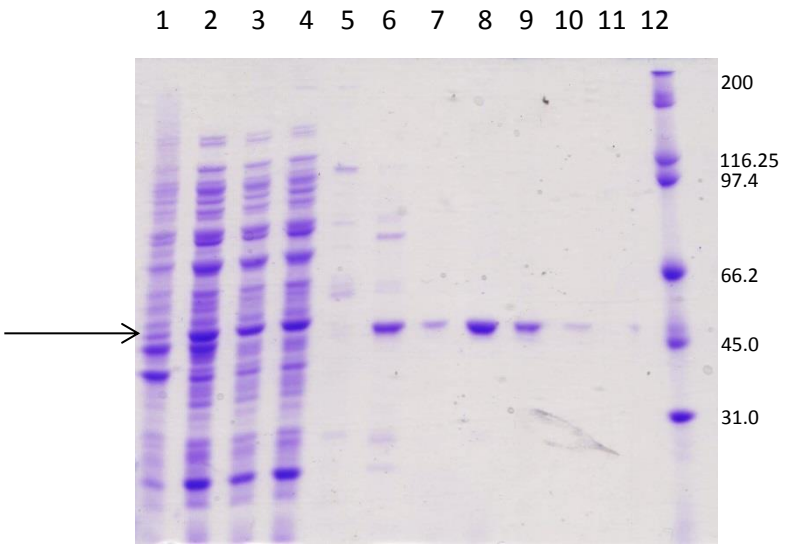


Figure 84 - SDS PAGE gel of ADH E. Lane 1: insoluble fraction, lane 2: soluble fraction, lane 3: 2nd flowthrough, lane 4: 0% His-Elute Buffer, lane 5: 1%, lane 6: 2.5%, lane 7: 5%, lane 8: 10%, lane 9: 30%, lane 10: 100% His-Elute Buffer, lane 12: Protein markers. The numbers on the right indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 45.4kDa.

3.2.1. SPECIFIC ACTIVITIES, SUBSTRATE SCOPE, CONVERSIONS

As performed previously, a broad range of substrates were tested with ADH E as part of a primary screen. Difficulty was noted with using raw enzyme containing fractions due to the extreme rate of reaction and radical variation in recorded activities principally due to the low volume of enzyme used in assays. Dilutions proved only moderately useful in combatting this issue, with discrepancies between expected activities from using one tenth of the enzyme, and actual activities recorded. It was not established whether the root cause of this issue was the small volume of raw enzyme needed, or if the reduction in enzyme concentration resulted in inherent storage stability issues. Either way, the problems were minimised before the primary screen.

Substrate	Specific Activity [U mg^{-1}]
Butanal	226.7
Decanal	8.6
2-Butanone	-
2-Decanone	-
3-Pentanone	-
Isobutanal	6.9
Cyclohexanone	-
Furfural	121.6
Acetophenone	-
Ethyl acetoacetate	-
Benzaldehyde	2.6
Pyruvate	-
Ethyl pyruvate	-
Ethyl levulinate	-
α-Tetralone	-
Levulinate	-
3-Methyl-2-butanone	-
2,3-Pentanedione	5.5
Ethyl 2-oxo-4-phenylbutyrate	-
4-Phenyl-2-butanone	-
1-Phenyl-1,2-propanedione	-
2,2,2-Trifluoroacetophenone	-
Ethyl 4-chloroacetoacetate	-
Ethyl 4,4,4-trifluoroacetoacetate	-
2',3',4',5',6'-Pentafluoroacetophenone	-

Table 18 - Primary screen for ADH E; substrates tested at 10mM concentration using NADH as cofactor.

Whilst initial assays indicated that ADH E could utilise either NADH or NADPH as cofactor, it had a very substantial preference for NADH, which was also more cost effective and therefore likely to be the preferred cofactor moving forward.

Very high activities were recorded with small aldehydes, as well as lower activities with furfural and benzaldehyde. Surprisingly, despite no activity with any other ketones, ADH E had some activity with 2,3-pentanedione. There is no obvious reason why this is the case, as no activity with halogenated substrates or ketoesters makes it less likely that electron density is particularly important for this enzyme.

Based upon these results, a secondary screen composed of a range of smaller aldehydes as well as lengthier ones for comparison indicated activity with standard aldehydes, but not where the substrate's carbon backbone is branched, as in the case of 3-methylbutanal. The previous belief that the substrate scope was independent of electron density was reinforced by a lack of activity with 4-chlorobenzaldehyde, yet benzaldehyde had a low activity. It may be the case that physical bulk was an issue though, as 5-norbornene-2-carboxaldehyde was equally ineffective as a substrate.

Substrate	Specific Activity [U mg^{-1}]
Ethanal	362.3
Propanal	376.4
Butanal*	225.3
3-Methylbutanal	-
Hexanal	99.0
Octanal	280.8
2,3-Butanedione	-
2,3-Hexanedione	3.5
2,3-Heptanedione	2.9
2,5-Hexanedione	-
3,4-Hexanedione	2.7
4-Chlorobenzaldehyde	-
5-Norbornene-2-carboxaldehyde	-
Glyoxal	-

Table 19 - Secondary screen with ADH E, Substrates tested at 10mM concentration using NADH as cofactor. * Repeat from Primary screen.

The apparent activity with diketones was rather surprising. Despite enzyme proportionality studies indicating high correlation between enzyme concentration and activity, the actual activity recorded remains very low and therefore on the limit of detection for the experimental method. Thus it is not beyond reasonable doubt that perhaps there is very low activity with 2,3-butanedione, but not detectable without considering other methods of study, more concentrated enzyme, or end point assays.

3.2.2. KINETICS

The substrates chosen for kinetic study were all short chain aldehydes, as these were the substrates with which ADH E had the greatest activity. In addition to substrates previously studied, methanal was also studied.

Ethanal and propanal have been plotted assuming that substrate inhibition does not exist, although it is certainly borderline in these cases. This is despite both methanal and butanal exhibiting clear signs of inhibition. As this enzyme was implicated in propanal degradation through its placement in the genome, it is hypothesised that propanal is likely to be the native substrate, and similar substrates are effectively inadvertent secondary substrates, i.e. the enzyme is promiscuous with respect to the length of substrate it accepts.

With methanal being a very harsh substrate to many enzymes, and butanal being longer than the anticipated native substrate, it is to be expected that they may exhibit signs of substrate inhibition. The inhibition noted in the case of butanal was particularly severe, with optimal activity being constrained to a very narrow range of concentrations.

As ADH E accepted both cofactors, both were tested as well, but all substrate kinetics was carried out with the preferred cofactor, NADH.

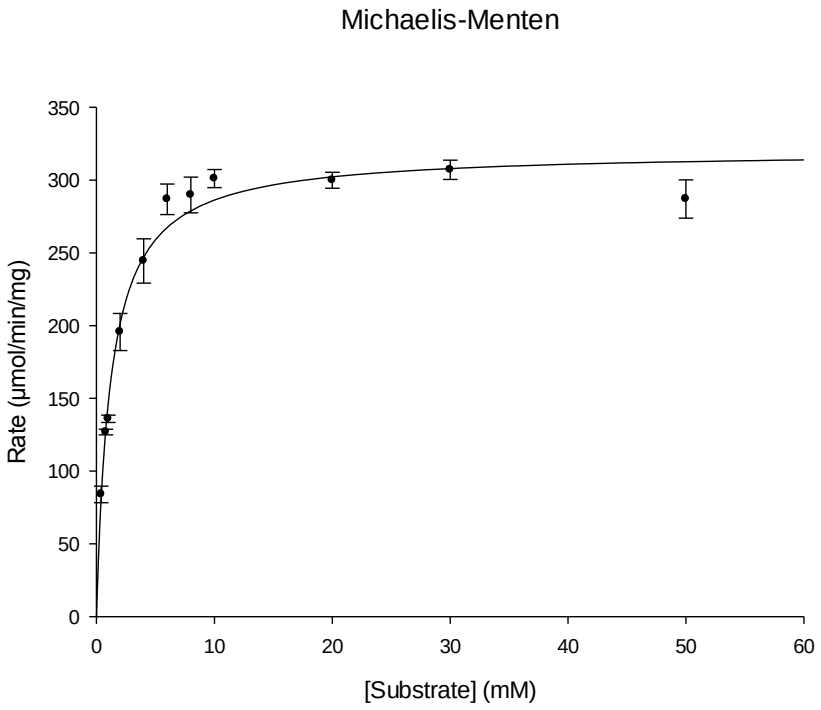


Figure 85 - Graph of rate of reaction vs [S] using ADH E and propanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Figure 86 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and propanal. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

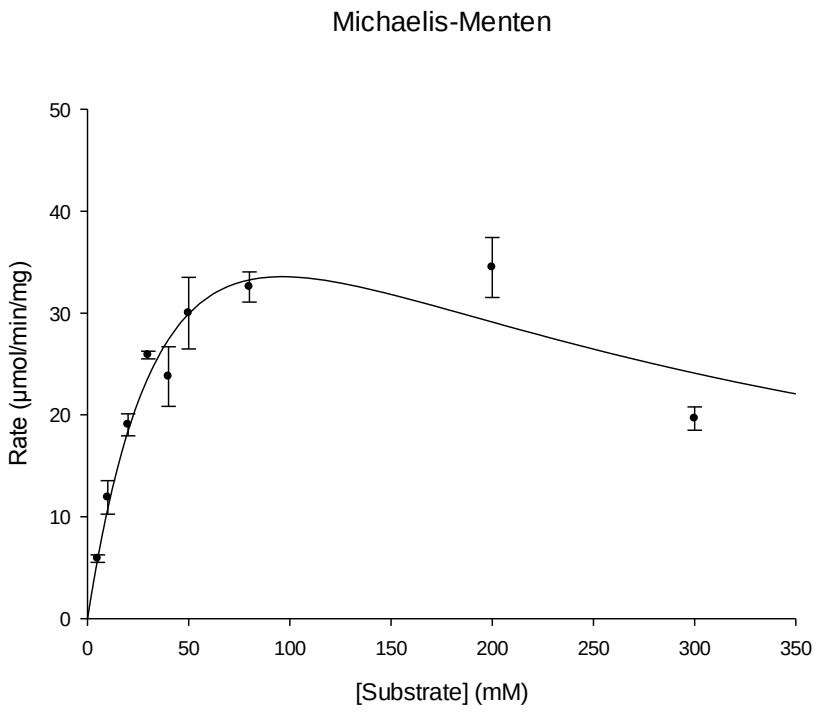


Figure 87 - Graph of rate of reaction vs [S] using ADH E and methanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

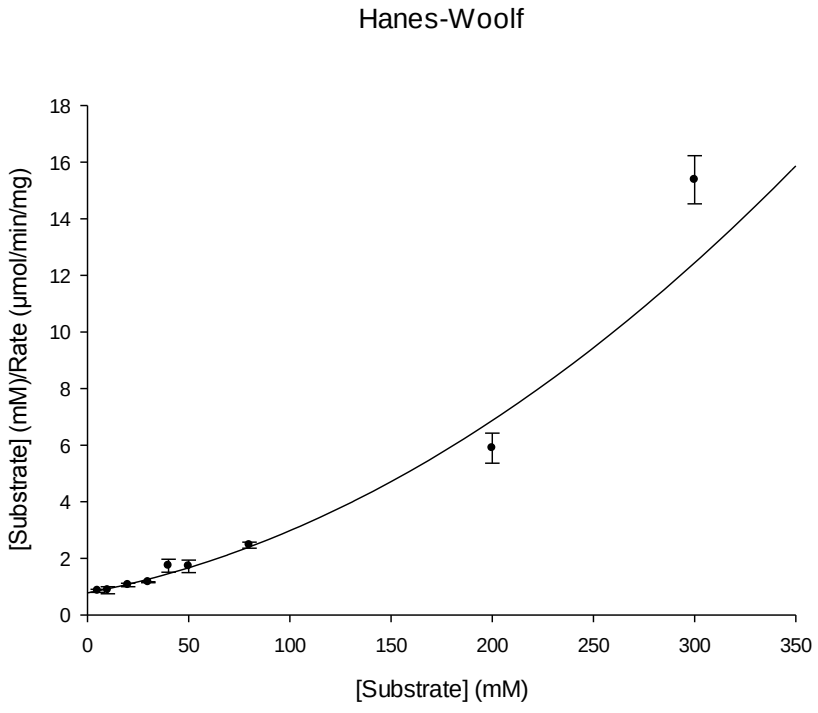


Figure 88 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and methanal, data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

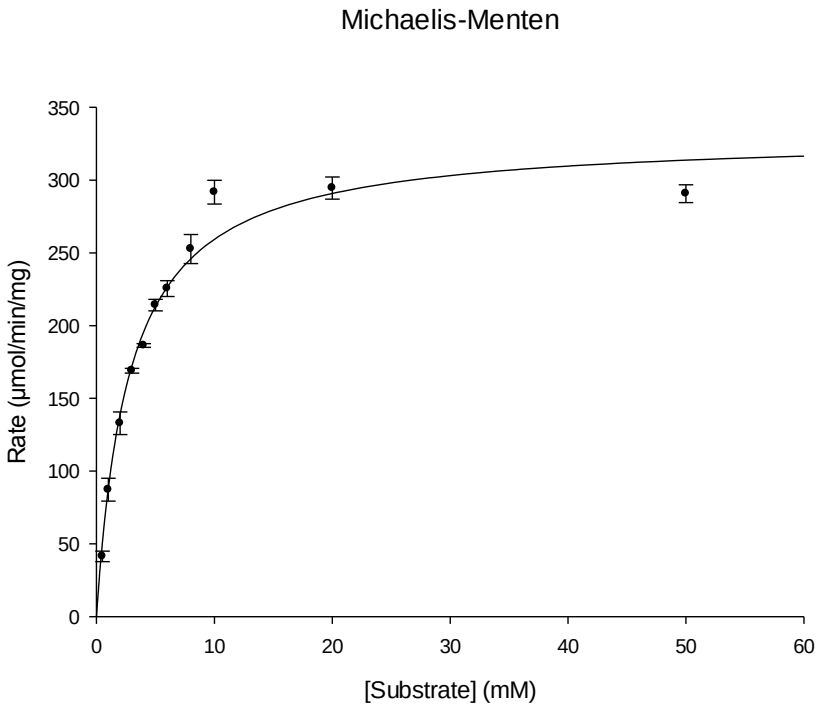


Figure 89 - Graph of rate of reaction vs [S] using ADH E and ethanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

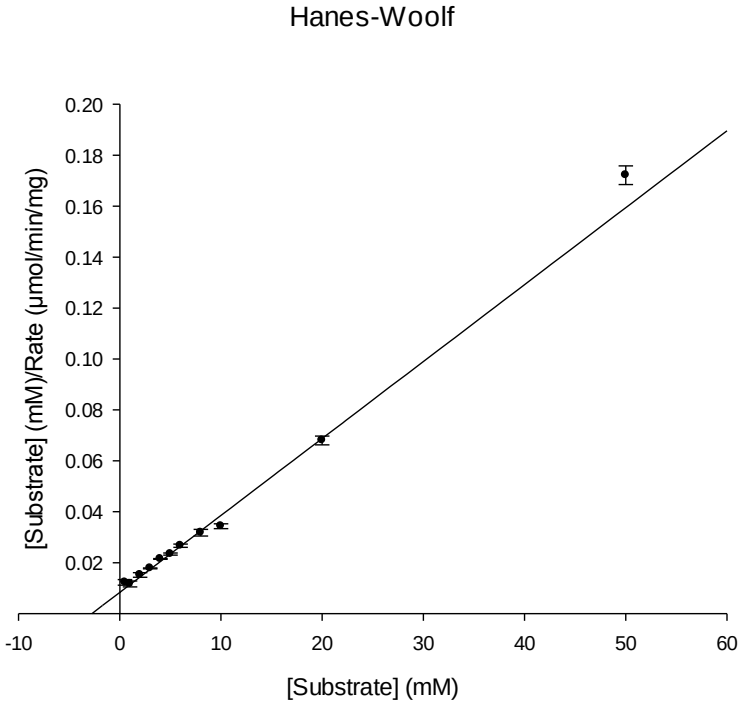


Figure 90 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and ethanal. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

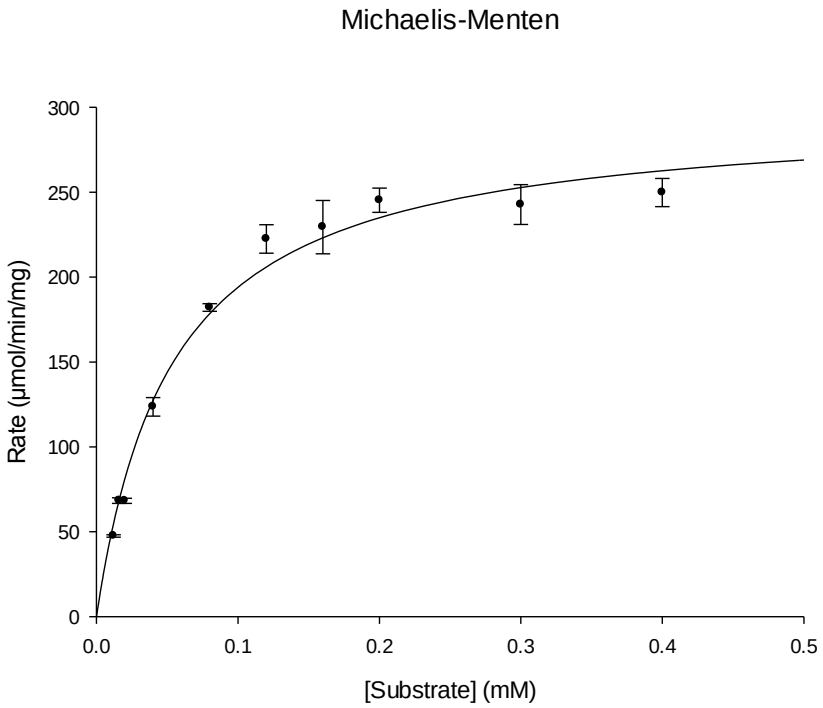


Figure 91 - Graph of rate of reaction vs [S] using ADH E and NADH. Data fitted to a Michaelis-Menten curve fitted using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM propanal, 50 °C, 1ml assay volume.

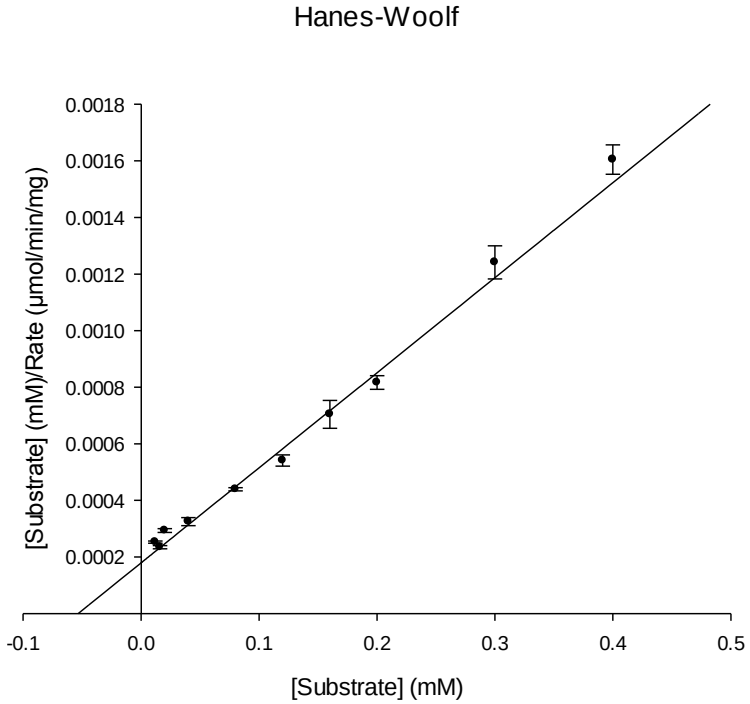


Figure 92- Hanes-Woolf plot ([S]/v vs [S]) for ADH E and NADH. Conditions: 50mM sodium phosphate pH 7.0, 10mM propanal, 50 °C, 1ml assay volume.

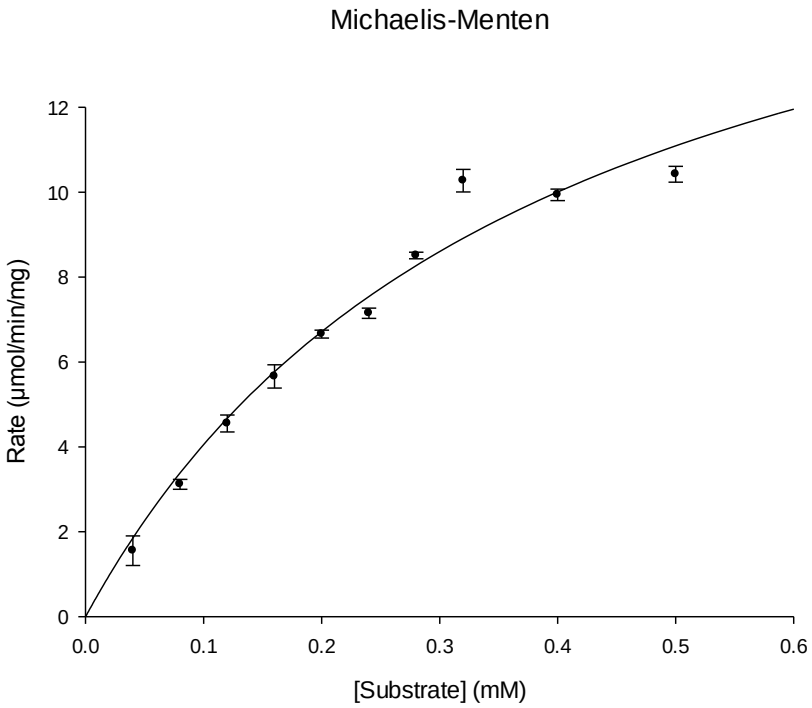


Figure 93 - Graph of rate of reaction vs [S] using ADH E and NADPH. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM propanal, 50 °C, 1ml assay volume.

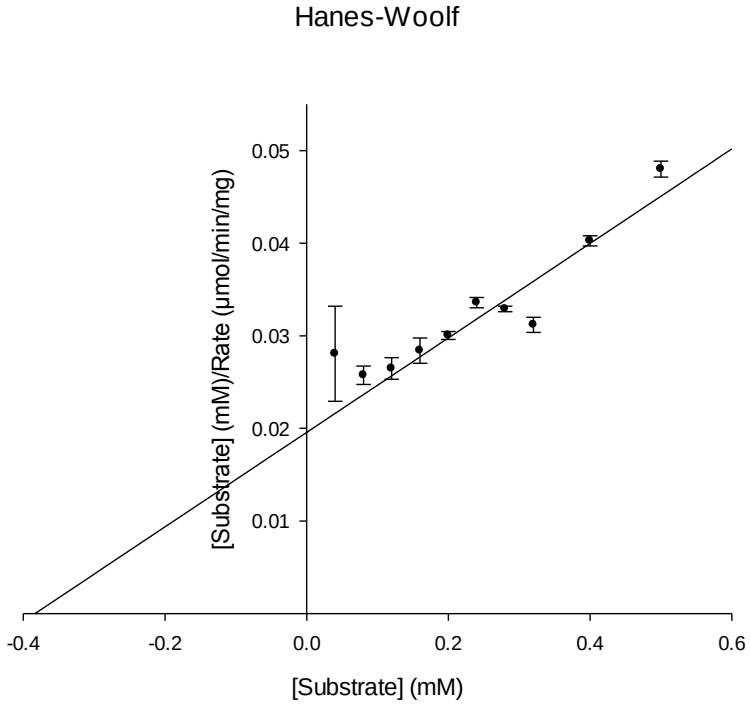


Figure 94 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and NADPH. Conditions: 50mM sodium phosphate pH 7.0, 10mM propanal, 50 °C, 1ml assay volume.

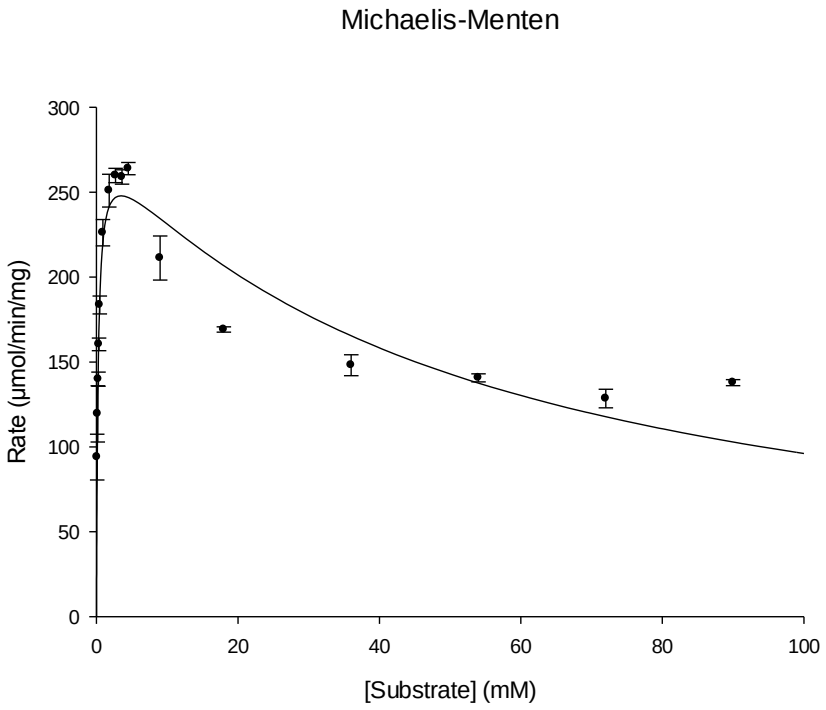


Figure 95 - Graph of rate of reaction vs [S] using ADH E and butanal. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

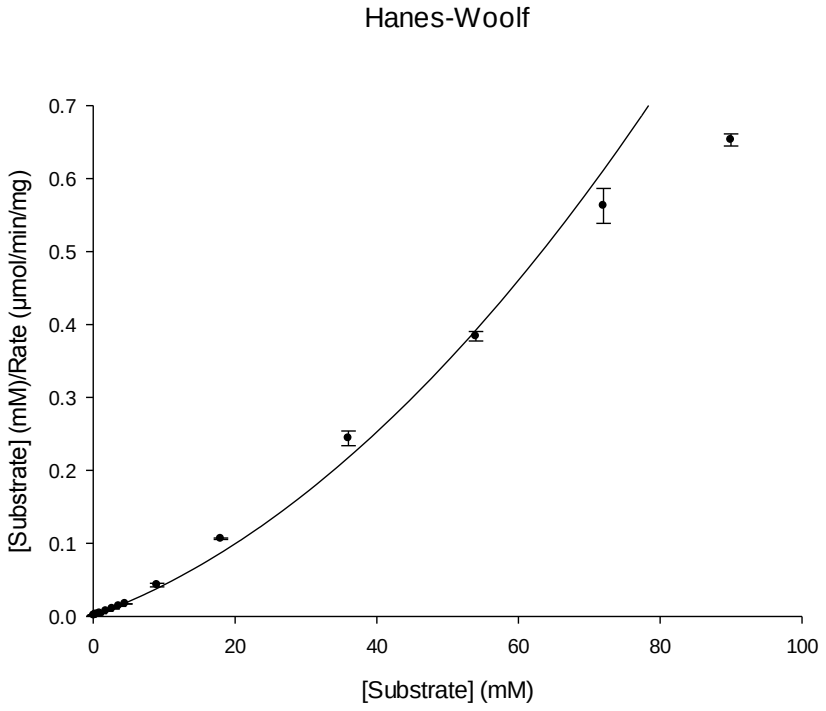


Figure 96 - Hanes-Woolf plot ([S]/v vs [S]) for ADH E and butanal. Data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
Propanal	320(±6)	1.18(±0.10)		15000	12000
Methanal	73.9(±24.0)	57.8(±28.8)	160(±89)	2400	58
Ethanal	331(±8)	2.76(±0.22)		15000	5400
NADH	298(±9)	0.05(±0.005)		14000	270000
NADPH	19.6(±1.62)	0.38(±0.06)		890	2300
Butanal	281(±9)	0.23(±0.03)	52.1(±6.1)	13000	55000

Table 20 - Summary of kinetic data with ADH E.

As expected, the kinetics indicated a very large bias towards using NADH as cofactor, two orders of magnitude difference in the specificity constant. Additionally, the best substrates tested were ethanal and propanal, if inhibition is discounted. However, relying solely on the specificity constant would imply that butanal is a more favoured substrate than any other tested.

This unusual circumstance could suggest that whilst ADH E is indeed part of an aldehyde degradation pathway, it has a preference for longer substrates, with a concomitant reduction in Michaelis constant, which is observable for methanal through butanal. With the exception of methanal, the V_{max} values obtained are close enough to be considered identical, definitely within their respective standard error values.

With reference to the specific activities recorded, it is apparent that either the ethanal and propanal V_{max} values are under-estimated, or the specific activities are over-estimated, as they do not match appropriately. Further work is needed to identify the root cause of this; however, the V_{max} is likely to be more accurate due to the number of readings from which it is based. More intriguingly though, is the hexanal and octanal values of 99 and 281 U mg^{-1} respectively. Although all specific activities are reported in the presence of 10mM substrate, not all substrates will dissolve to that concentration. Whereas in kinetics experiments solvents are used to aid dissolution of substrates, they are avoided almost entirely in screening to prevent solvents from affecting the rates recorded; occasional exceptions exist for solid substrates. Liquid substrates are added in a pure form, and sufficiently mixed to ensure a saturating concentration is present.

As the saturating concentration of most substrates is not known in the buffers used, using pure water as substitute is viable in some cases, but not all. Using existing solubility studies by way of example, it is also important to note that the solubility of many aldehydes can reach a minimum at the assay temperature of 50^{o}C, which is a regrettable but unavoidable circumstance.

With reference to the butanal kinetics and to the hexanal and octanal specific activities, it seems apparent that ADH E has approximately the same V_{max} for aldehydes ethanal through at least to octanal, but probably higher aldehydes as well. Therefore it could be predicted that K_M and K_i values will drop fairly substantially the larger the substrate gets, leading to the situation whereby ADH E will be incredibly active with larger substrates, but only at minute concentrations. The decanal specific activity may indicate that by that length of substrate this no longer occurs, or that by that point there is insufficient substrate for a true reading to be observed. Further work, particularly additional kinetics with longer substrates, will need to be carried out to substantiate these predictions.

Given that branching of the carbon chain results in lack of activity, and cyclic and aromatic substrates are generally ignored, then it appears that ADH E is restricted to linear aldehydes

only, excepting furfural and incidental activity with benzaldehyde. If the diketone activity is in fact true activity, then it would follow that only linear diketones would be accepted. That said, the rate obtained from these diketones is very low, although it would be of interest to determine their kinetic constants to see if they follow the same pattern as the aldehydes. Further work with furfural would also be useful, as this substrate is of particular interest as it acts as a toxin in the biodegradation of lignocelluosic feedstocks (Chang and Yao, 2011; Ji et al., 2011; Joshi et al., 2011).

3.2.3. STABILITY ANALYSIS

Initial thermostability analysis considered only two timepoints, 5 minutes and one hour and with two temperatures, 70 and 80 ^{o}C. Testing with 10mM furfural indicated 61% and 31% activity was retained after five minutes incubation at 70 and 80 ^{o}C respectively. None was detected after one hour at either temperature.

Preliminary solvent work was similar, only testing a pair of solvents at relatively low concentrations. The results are tabulated below.

Volume (vol%)	Activity (DMSO) [%]	Activity (Ethyl acetate) [%]	Activity (Acetonitrile) [%]
1	70	85	60
5	37	89	35
10	39	83	n.d.

Table 21 - Solvent stability analysis of ADH E, n.d. indicates not determined

On the basis of these assays, ethyl acetate was chosen as solvent for stock solutions of substrate where it was necessary.

An in-depth investigation into the stability of this enzyme was not carried out due to the limited substrate scope, particularly with regards to ketones. However, on the basis of results obtained it appears that this enzyme is somewhat less resistant to solvents and temperature when compared to others studied.

3.2.4. OPTIMAL TEMPERATURE AND MELTING TEMPERATURE

One area that was examined was the optimal temperature, which had to be carried out with butanal rather a smaller aldehyde due to the boiling point of ethanal and propanal. Comparing the Michaelis-Menten curve obtained previously and the activity noted at 50 ^{o}C in the optimal temperature study, it is obvious that substantial activity was lost. This is believed to be due to the use of a diluted enzyme fraction to reduce the observed rate of reaction to recordable levels.

That said, the temperature curve indicates that the optimal temperature is likely either 70 or 75 ^{o}C, which is approximately where expected; although ADH E has no long term stability at such high temperatures, it certainly will be able to survive sufficiently for initial rates to be determined. As there is little reduction in activity at 75 ^{o}C, it would be preferable to repeat the thermostability test at higher temperatures, 80 or 85 ^{o}C to ensure conclusively that this is the optimal temperature. Alternative substrates would have to be used, due to the boiling point of butanal at 74 ^{o}C.

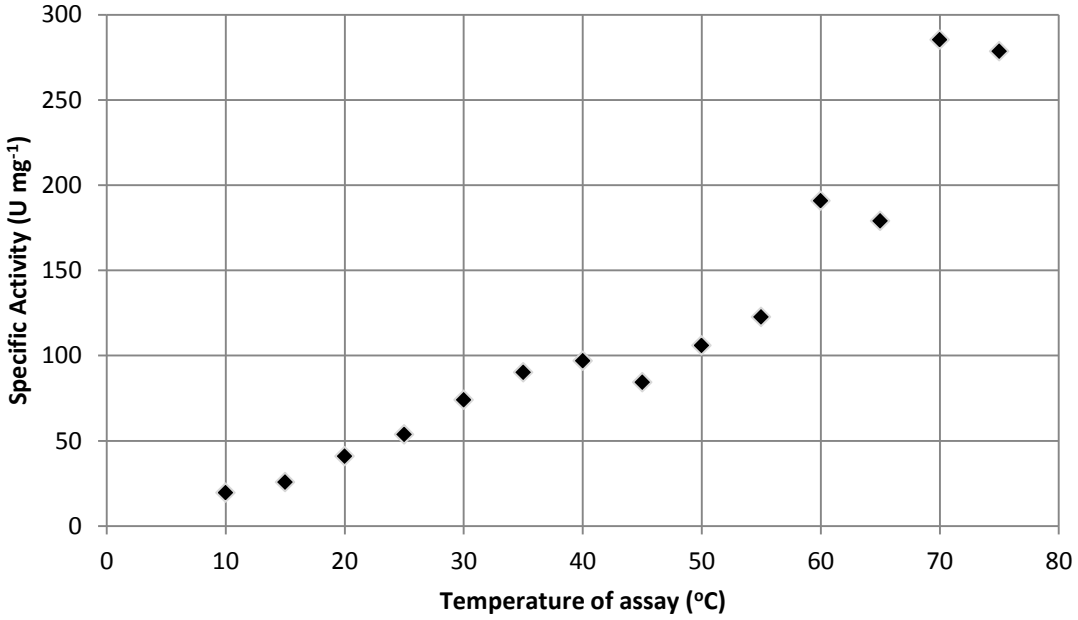


Figure 97 - Optimal Temperature study of ADH E, using NADH as cofactor and 10mM Butanal as substrate.

The thermal shift assay provided an inflection point at 80.3 °C, the highest point of all enzymes tested, excluding the second point of ADH D. The longer term thermostability of this enzyme would contrast with its apparent high T_m, as the brief testing done indicated a half-life of minutes at 70 °C and seconds or less at 80 °C. This is an area that should be examined in greater detail in the future to elucidate the useful thermostability of this enzyme for industrial purposes.

3.2.5. CHEMICAL ANALYSIS

Due to the enzyme accepting both cofactors, at least to a limited degree, assays were run with both cofactors and then analysed to see if there was a difference over a longer time period than the initial rates.

As envisaged, activity was noted with butanal, furfural and decanal, at least initially. The overnight assay did not produce a detectable quantity of decanol which is concerning. Similarly, low activity was noted with 4-chlorobenzaldehyde in the one hour assay, but not in the 16 hour assay. This is particularly interesting as the previous work did not indicate any activity with this substrate.

Whilst quantitative analysis was occasionally variable, presence or absence of product is relatively easy to detect by GC/MS. It is also worth noting that ethyl acetate was used where necessary to produce stock solutions of substrates such as decanal which are relatively insoluble.

	Conversion			
Substrate	**1 Hour**		**16 Hours**	
	NADH	**NADPH**	**NADH**	**NADPH**
5-Norbornene-2-carboxaldehyde	N/A	N/A	0%	0%
Butanal*	32%	11%	27%	2%
Furfural	Activity	Activity	28%	1%
Methylbenzoyl formate	-	-	-	-
Decanal	2%	0%	0%	0%
Benzaldehyde	0%	0%	0%	0%
4-Chlorobenzaldehyde	2%	0%	0%	0%
2,3-Pentanedione	N/A	N/A	0%	0%
2,3-Butanedione	N/A	N/A	0%	0%

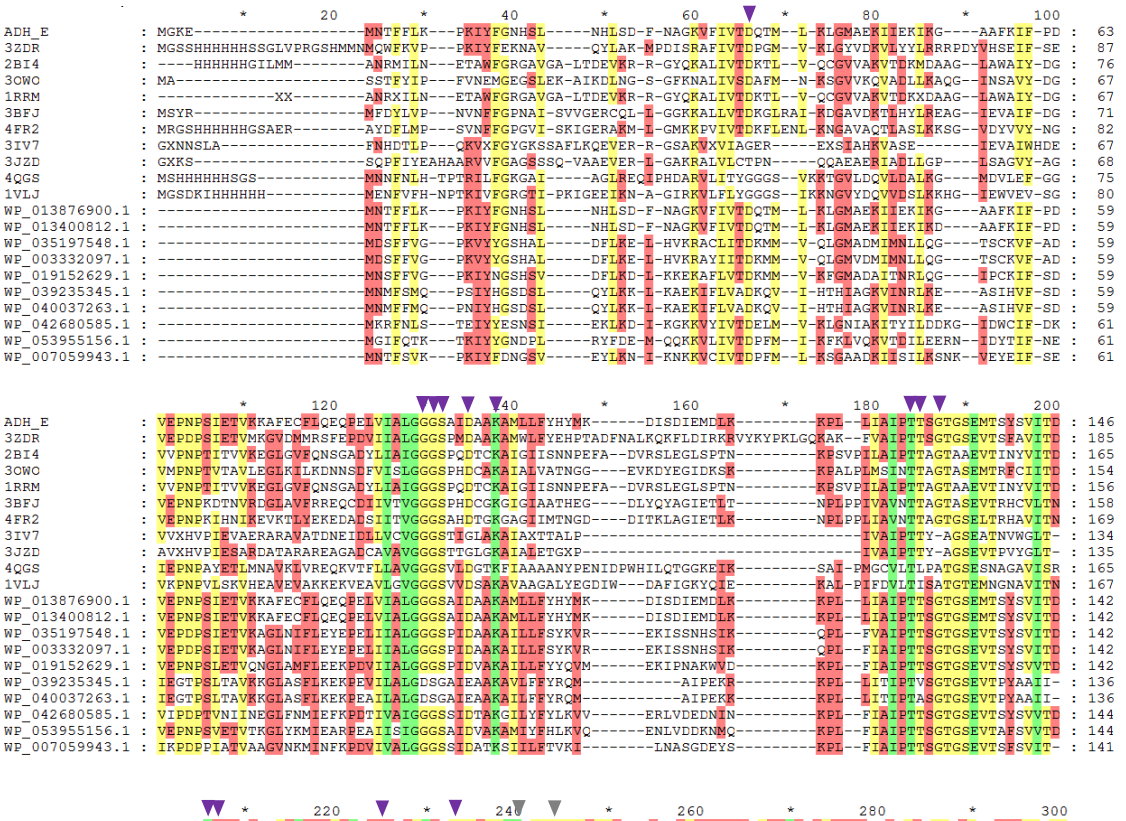
Table 22 - Summary of GC/MS data obtained with ADH E. N/A indicates not tested, "-" indicates substrate unsuitable for analysis. *Evaporation suspected with this substrate.

In particular, the entries for benzaldehyde and decanal suffered from poor signal detection on the 16 hour assay, which may have obfuscated some activity in the case of decanal and NADH.

Finally, it appears that activity noted with diketones was in fact illusionary, perhaps benzaldehyde too. Previously noted as on the limit of detection, it may be the case that in this particular instance that the change in absorbance was incidental or unrelated to real enzymatic activity.

3.2.6. SEQUENCE ALIGNMENT

The MSA conducted for ADH E was slightly different to others carried out due to the lack of protein sequences identified through the NCBI Seqr Sequence Search as this only provides sequences with >70% identity to the sequence provided. Thus the ten closest sequences were obtained using a standard BLAST search using the Reference Sequence database. Compared to all other ADHs, ADH E was unique in having very few (<10) sequences with >70% sequence identity.



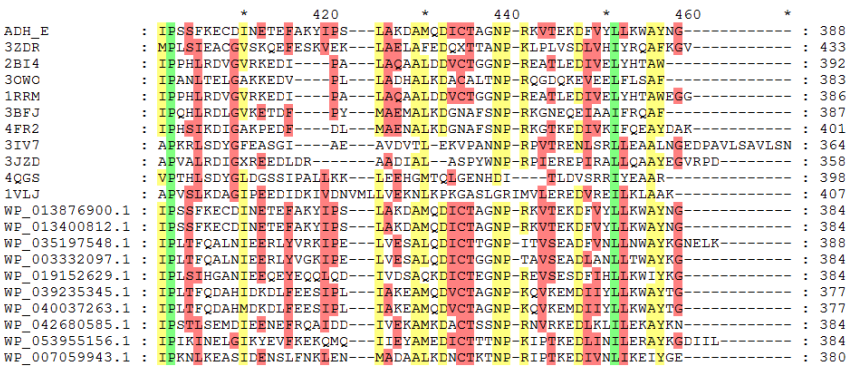


Figure 98 - Multiple Sequence Alignment of ADH E with most similar protein sequences associated with crystal structures and obtained from the PDB and most similar protein sequences obtained from the NCBI RefSeq Database. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to crystal structures from the PDB or proteins from the NCBI RefSeq Database. Proteins are annotated as ADHs from the genera *Geobacillus, Bacillus, Anaerosalibacter* and *Clostridium*. Crystal structures are annotated as follows: 3ZDR is a the ADH domain of a bifunctional ADHE dehydrogenase from *G. thermoglucosidasius*, 2BI4 is a lactaldehyde:1,2-propanediol oxidoreductase from *E. coli*, 3OWO is an iron dependent ADH from *Zymomonas mobilis*, 1RRM is a lactaldehyde reductase from *E. coli*, 3BFJ is a 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae*, 4FR2 is an ADH from *Oenococcus oeni*, 3IV7 is an iron containing ADH from *Cornyebacterium glutamicum*, 3JZD is a putative ADH from *Ralstonia eutrophia*, 4QRS is the aldehyde reductase YqhD from *E. coli*, 1VLJ is a butanol dehydrogenase from *Thermotoga maritima*. Purple triangles indicate the putative active site residues and grey triangles indicate the metal binding site. All these sites provided via the NCBI Conserved Domain Database and obtained through the Conserved Domain Search. Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997).

With no highly similar proteins in the NCBI database, there is significantly less conservation among the sequences aligned, with very few residues conserved among all proteins in the MSA. In particular, aside from the first two entries, WP_013876900.1 and WP_013400812.1, which are from *Geobacillus* species and are near identical, the next sequences have 63% identity, indicating that ADH E is relatively unusual and notably different from any other ADH studied.

Another point of interest is that the nearest crystal structure is 3ZDR, or ADH A. As ADH A was also studied as part of this work, its substrate scope is well known, and restricted to linear aldehydes. The metal co-ordination site is D225, H229, H294, H308, which correspond to D240, H244, H314, H328. With one exception at one site, all these are conserved in ADH E and in all sequences in the MSA. Cofactor specificity appears to be controlled by D57 (D66) which clashes with the phosphate group of NADP(H) (Extance, 2012). This site is also annotated as part of the active site.

2BI4 is an NADH dependent lactaldehyde reductase from *E. coli,* also known as FucO. It is a dimer and the first reported structure for an iron-containing ADH that has been shown to contain iron. Although 2BI4 has a greater affinity for zinc than iron, zinc severely inhibits the enzyme. Similar to 3ZDR, the metal co-ordination site is identified and the cofactor specificity is dependent on D39 (D66). The NAD^+ binding site was experimentally confirmed to be GGS96-98 (GGS130-132) by considering the mutant G97E (G131E) which had no detectible activity or NAD^+ binding. A second mutant D39G (D61G) resulted in 2BI4 being able to utilise NADPH and NADH, with affinity roughly equal. There was a substantial drop in k_{cat} when using NADPH as cofactor though; 104 s^{-1} when using NADH and 31 s^{-1} when using NADPH. It is believed that K40, T41, Q44 (K62, T63, Q66) may provide a pocket for the phosphate group to bind and it is

solely the D39 (D61) residue that hinders (Montella et al., 2005). Given that these residues are similar to ADH E, which already has activity with NADPH, the same mutation may allow for similar results. 2BI4 is a highly stereospecific enzyme, with preference for *S*-1,2-propanediol, and also regiospecific for the oxidation of primary alcohols. It has a limited substrate scope, able to catalyse primary alcohols, diols and triols, but only aliphatic examples. Rates of reaction are also very poor, with the highest k_{cat} of 4 s^{-1} with 1,2-ethanediol (K_M of 51mM) (Blikstad and Widersten, 2010).

3OWO is the apo form of a dimeric, NADH dependent, ADH from *Zymomonas mobilis*. It is a member of the iron-containing family of ADHs, oxygen sensitive, and believed to be a key enzyme in the cytosolic respiratory system, converting ethanal to ethanol and providing necessary NADH. As may be expected given the conservation of the residues, the same motif is indicated as the cofactor binding site and D39 (D61) is again considered the principal reason 3OWO is NADH dependent and not NADPH dependent. Key residues for substrate binding are believed to be F149 and F254 (F195 and F302), which sterically regulate the substrate molecule, in this case ethanol (Moon et al., 2011). These residues are conserved in ADH E, and therefore may be relevant with regards to the substrate scope of this enzyme.

1RRM is a lactaldehyde reductase from *E. coli*, but has not been published.

3BFJ is a decameric 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae*. It is NAD dependent, iron-containing, and is proposed as a key enzyme in the use of glycerol. As with previous enzymes, three histidine residues and one aspartic acid residue are metal co-ordinating. The substantial difference with this enzyme is the decameric quaternary structure, in which there are five dimers. Two interactions mediate this, the first is ionic between D322 and R333 (D383 and R394), the second is a hydrogen bond between N314 and G317 (N375 and G378). None of these four residues are conserved in ADH E, but are typically conserved in 1,3-propanediol dehydrogenases (Marçal et al., 2009).

Lastly, 4FR2 is a dimeric NADH dependent ADH From *Oenococcus oeni*. Although a member of the iron-containing ADH family, the active enzyme contains nickel and is entirely inhibited by zinc. It is the first crystal structure from a gram-positive bacteria and can catalyse a range of small aldehydes and alcohols including 1,2-propanediol, 1,3-propanediol, glycerol, 2-propanol and butanal. Identical residues to those previously noted have been identified as key for cofactor specificity and binding (Elleuche et al., 2013).

In conclusion, ADH E is an unusual enzyme, with few sequences with >70% sequence identity on record. The most similar enzymes are either 1,2-propanediol dehydrogenases or 1,3-propanediol dehydrogenases, typically with the ability to accept varied small substrates such as glycerol and butanol, but with very weak rates of reaction and high Michaelis constants. ADH E has a turnover number between ten and one hundred times higher than similar enzymes noted above, each with their most preferred substrates and cofactors. Given ADH E's observed activity with larger substrates, including octanal and furfural, it seems likely that it is a different ADH to those detailed above. Speculatively, the change in substrate scope could be mediated by residues annotated as part of the active site and are not conserved such as I207 and L225.

Despite this, the conserved residues and experimental evidence obtained with other enzymes imply it may be possible to drastically alter the cofactor specificity with a single mutation, D61G, perhaps allowing substantial activity with both cofactors simultaneously. This would be a very significant advance as it would allow ADH E to become a more useful thermophilic enzyme for cofactor regeneration. More tentatively, there may be opportunity to remove

NADPH activity entirely and increase NADH activity by manipulating residues believed to assist in NADPH binding.

It would be interesting to determine whether a metal ion is used by ADH E, given the range of metal ions that utilised by related enzymes, it may be instructive to examine this area. Another question that arises from the MSA is if the sequence for ADH E is entirely accurate. It is possible that an additional four amino acids have been appended to the N-terminus, but this is not entirely clear. As an N-terminal His tag has been used in this investigation it is highly unlikely to have affected the enzyme beyond any effect the tag has already caused, but it is worth noting regardless.

3.2.7. SUMMARY

ADH E is an alcohol dehydrogenase with very high activity with short chain aldehydes. Due to this particular substrate scope detailed investigations were not carried out; however, with substantial potential in the detoxification of furfural and in bioremediation of short-chain aldehydes, this enzyme has significant capability. Additional benefit may be gained as part of a two enzyme, two substrate cofactor regeneration system. Due to the enzyme's ability to utilise both cofactors, ADH E demonstrates usefulness as a multipurpose regeneration system with high proficiency at producing the necessary reduced cofactors for a variety of industrial and research purposes.

Although a thermophilic enzyme, it is not especially resistant to high temperatures over long timescales, but demonstrates short-term resistance to temperatures higher than would be expected given the growing temperature of *G. thermoglucosidasius*. This is reflected in variable solvent resistance too, with apparent immunity to ethyl acetate at low concentrations, suitable for cosolvent purposes but sensitive to both acetonitrile and DMSO.

With regards to native activity, it seems highly likely that ADH E is involved in propanal reduction yet may equally be involved in breaking down a variety of minimally soluble higher aldehydes. Speculatively, these aldehydes may interfere with the cell membrane and therefore may need to be converted to alcohols to prevent damage.

3.3. ADH F

The first enzyme to be tested from a second group of putative ADHs, ADH F expressed and generally purified without issue, as shown below in a SDS-PAGE gel. On occasion, however, expression was substantial enough to overwhelm the column used for purification, resulting in infrequent minor contaminants, which would counter-intuitively appear at greater concentrations of imidazole in the elution buffer. This was attributed to these contaminating proteins being "locked in" to the column by ADH F, by its surrounding of the column. Therefore in early elutions some ADH F washed off, exposing the other proteins to the imidazole containing buffer, and eluting later than would otherwise be expected.

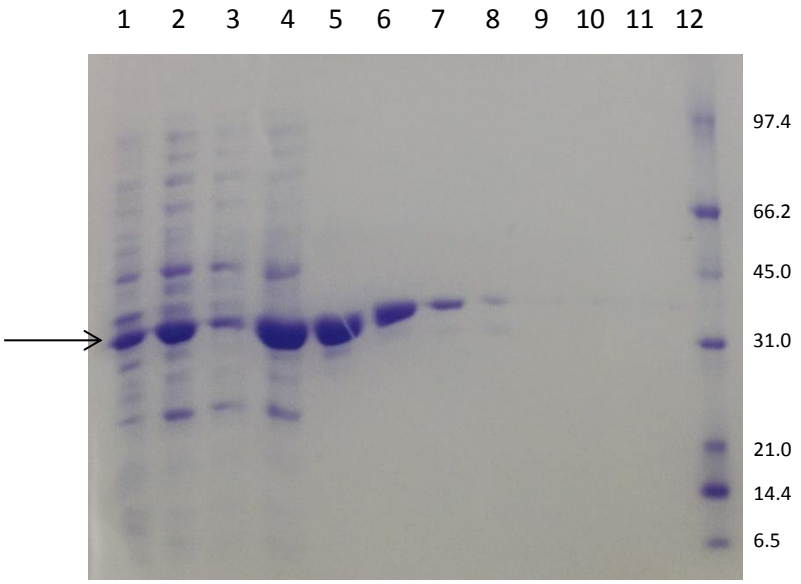


Figure 99 - SDS PAGE gel of ADH F. Lane 1: insoluble fraction, lane 2: soluble fraction, lane 3: 2nd flowthrough, lane 4: 0% His-Elute Buffer, lane 5: 1%, lane 6: 2.5%, lane 7: 5%, lane 8: 10%, lane 9: 30%, lane 10: 100% His-Elute Buffer, lane 12: protein markers. The numbers on the right indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 34kDa.

3.3.1. SPECIFIC ACTIVITIES, SUBSTRATE SCOPE, CONVERSIONS

As is customary, a primary substrate screen was conducted using the same substrates as with previous enzymes. Similar to ADH D, a wide range of substrates were accepted by this enzyme. Unique among all enzymes tested though, was the apparent ability of ADH F to utilise both cofactors to a similar degree, and thus the primary screen was conducted twice, once with each cofactor.

Substrate	Specific Activity (U mg^{-1}) [NADH]	Specific Activity (U mg^{-1}) [NADPH]
Butanal	1.10	5.45
Decanal	0.33	1.52
2-Butanone	-	-
2-Decanone	-	-
3-Pentanone	-	-
Isobutanal	1.18	5.29
Cyclohexanone	-	0.91
Furfural	3.14	8.84
Acetophenone	-	0.33
Ethyl acetoacetate	0.12	1.31
Benzaldehyde	4.75	8.35
Pyruvate	-	0.14
Ethyl pyruvate	6.36	3.42
Ethyl levulinate	-	0.12
Levulinate	-	-
α-Tetralone	-	-
3-Methyl-2-butanone	-	0.07
2,3-Pentanedione	16.52	10.75
Ethyl 2-oxo-4-phenylbutyrate	1.60	3.94
4-Phenyl-2-butanone	-	0.27
1-Phenyl-1,2-propanedione	20.74	11.45
2,2,2-Trifluoroacetophenone	2.51	9.44
Ethyl 4-chloroacetoacetate	5.18	1.84
Ethyl 4,4,4-trifluoroacetoacetate	0.69	2.83
2',3',4',5',6'-Pentafluoroacetophenone	0.30	2.84

Table 23 - Primary screen of ADH F with both cofactors.

It was not clear from these data which cofactor was preferred by ADH F, as although the majority of rates were higher with NADPH, there were substantial improvements in a few cases with NADH, such as 1-phenyl-1,2-propanedione and ethyl 4-chloroacetoacetate.

Also notable is that with NADPH, ADH F lacked activity with very few substrates in the screen, specifically monoketones, tetralone and levulinate. Based around the tendency of this enzyme to prefer diketones, a variety were collected for the secondary screen, along with several aldehydes and related industrially relevant substrates such as halogenated acetophenones. Structurally similar substrates such as methyl acetoacetate were also added to see if activity was limited to specific substrates.

Substrate	Specific Activity (U mg^{-1}) [NADH]	Specific Activity (U mg^{-1}) [NADPH]
Methanal	0.14	-
Propanal	0.65	-
Hexanal	5.10	0.60
Octanal	6.69	2.28
2,3-Butanedione	15.86	24.36
2,5-Hexanedione	-	-
2,3-Hexanedione	19.82	21.80
2,3-Heptanedione	21.13	32.80
3,4-Hexanedione	8.29	1.50
3-Methylbutanal	-	-
Cycloheptanone	-	-
2,2',4'-Trichloroacetophenone*	2.38	-
4'-Chloroacetophenone	0.18	-
4'-Bromoacetophenone	0.37	-
2-Chloroacetophenone	0.24	-
Methyl acetoacetate	0.56	-
Ethyl 2-chloroacetoacetate	11.85	11.15
4-Chlorobenzaldehyde	5.23	3.97
Glyoxal	12.38	13.83
Propiophenone	-	-
5-Norbornene-2-carboxaldehyde	9.97	6.65
Benzoin	-	-
Benzil*	0.27	-
Benzophenone*	-	-

Table 24 - Secondary screen of ADH F with both cofactors. *These substrates tested at 1mM concentration, from 100mM stock solution prepared in DMSO.

This enzyme demonstrated activity with almost every substrate provided to it. Limitation in size is only apparent with benzophenone, although perhaps even this may be averted with higher concentrations of substrate. All solid substrates were prepared in DMSO for test, and due to poor solubility even then some were tested at 1mM concentration.

Branched aldehydes and other excessively large substrates, specifically 3-methylbutanal and propiophenone, were also unsuitable for ADH F. Furthermore, it is shown that diketones must be adjacent, as 2,5-hexanone was equally unsuitable. Aside from this, there was noticeable activity with every substrate, with at least one cofactor, if not both.

One particular quirk of ADH F was with butanal, using NADPH as cofactor. The activity reported only occurred after a substantial lag phase, where little or no activity was recorded. This was not attributable to product activation, as premixing 1-butanol with the enzyme had no effect upon assays. Additionally, this behaviour was not noted when using any other aldehyde or NADH as cofactor.

Overall though, based solely upon the screening, ADH F demonstrated an extensive substrate scope, including all industrially relevant substrates tested and analogues of commercially useful structures. The ability to use NADH as well as NADPH and reasonable rates of activity with many substrates of interest were also noted. Therefore, ADH F became a high priority for further investigation.

3.3.2. KINETICS

As with the substrate screening, kinetics were carried out in duplicate, with both NADH and NADPH used as cofactors for each substrate tested. For this reason only five substrates were considered, with each cofactor being examined in addition to this. Contrasting with other enzymes studied, the majority of substrates used for kinetics work had poor solubility in buffer and therefore stock solutions were prepared in solvent. In this case acetonitrile was used.

Noted previously, there is often a non-existent operating range for substrate concentration when analysing the kinetics using substrates that are virtually insoluble in buffers. Solvent concentrations were carefully monitored in accordance with previously determined resistance studies; however, in some cases it was impossible to avoid using higher than desired solvent concentrations in order to provide enough substrate to the assay. These cases are where the substrate concentration exceeded 20mM, which corresponds to a 2vol% solvent concentration in the assay. At the other end of the spectrum, to accommodate particularly low substrate concentrations, e.g. 1-phenyl-1,2-propanedione, which necessitated having concentrations as low as 20µM, the original stock solution was diluted.

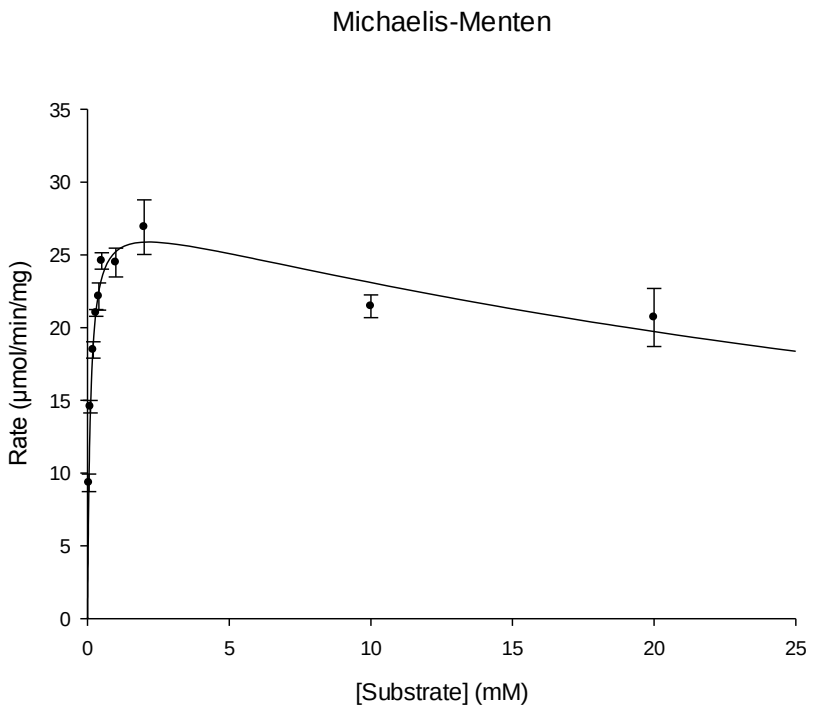


Figure 100 - Graph of rate of reaction vs [S] using ADH F and 1-phenyl-1,2-propanedione in acetonitrile. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

Figure 101 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 1-phenyl-1,2-propanedione in acetonitrile, data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

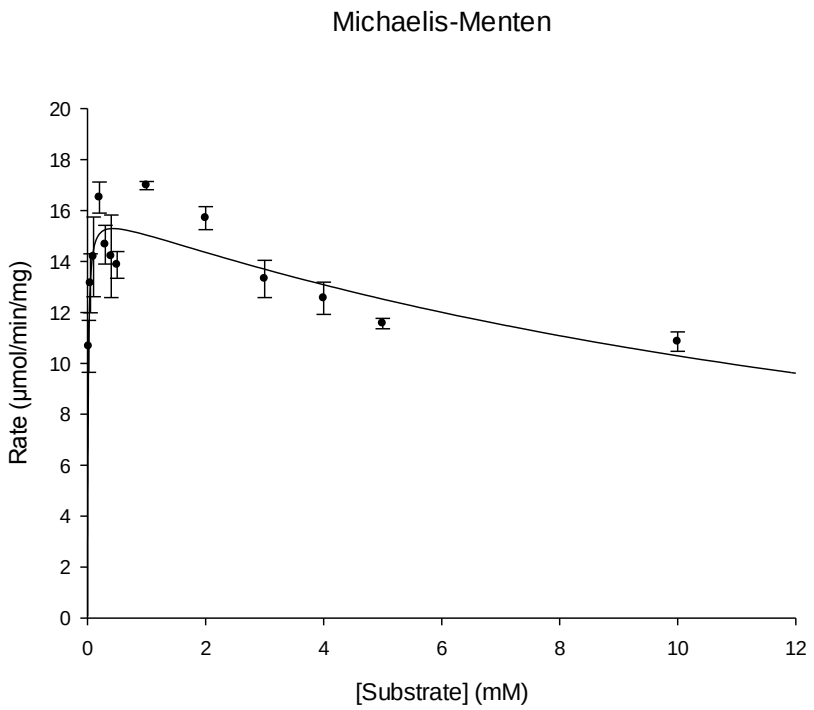


Figure 102 - Graph of rate of reaction vs [S] using ADH F and 1-phenyl-1,2-propanedione in acetonitrile. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Figure 103 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 1-phenyl-1,2-propanedione in acetonitrile, data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

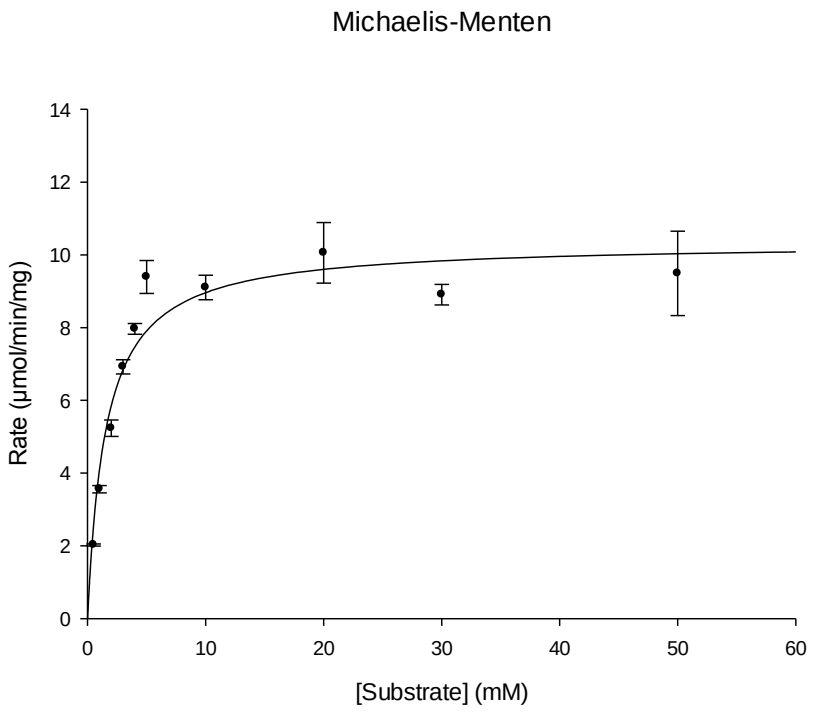


Figure 104 - Graph of rate of reaction vs [S] using ADH F and furfural. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

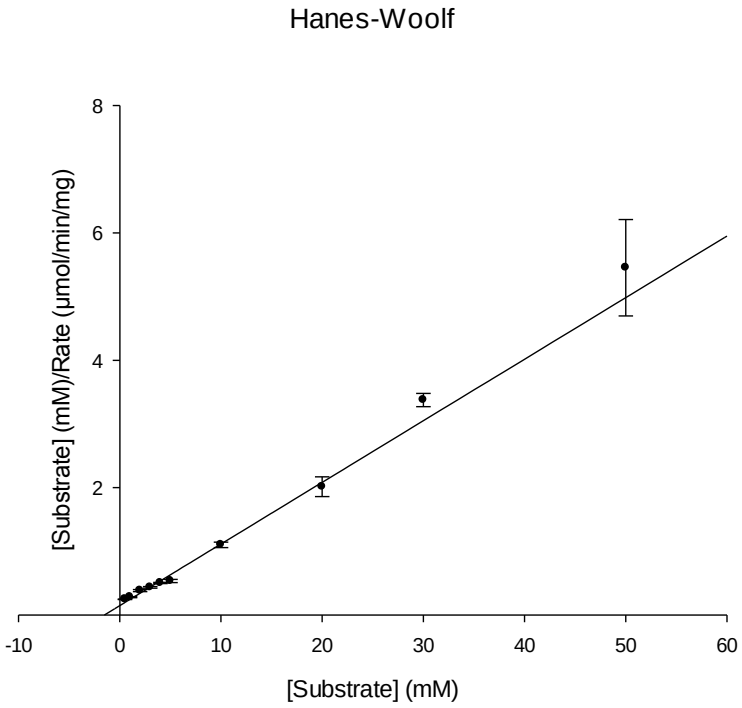


Figure 105 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and furfural. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

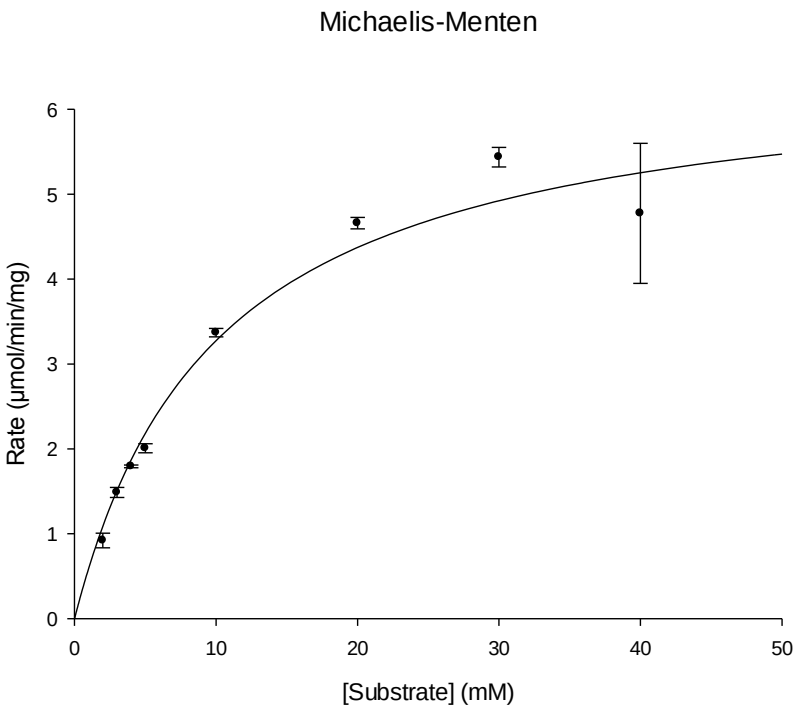


Figure 106 - Graph of rate of reaction vs [S] using ADH F and furfural. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

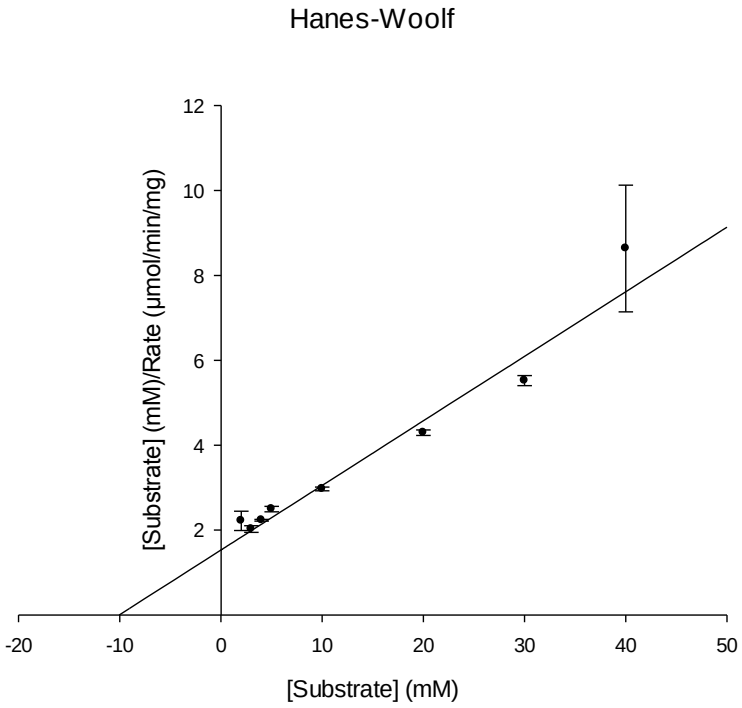


Figure 107 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and furfural. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

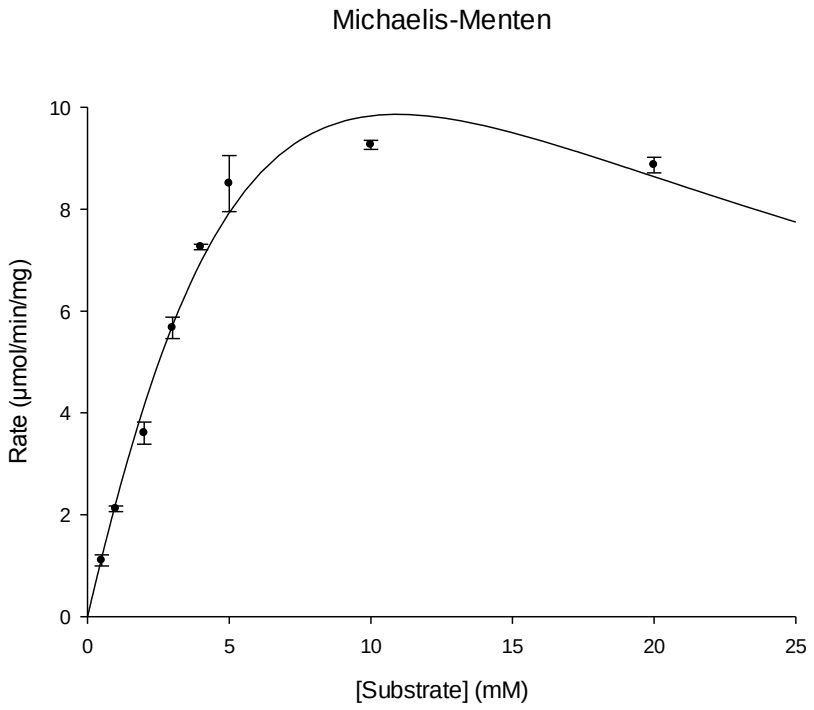


Figure 108 - Graph of rate of reaction vs [S] using ADH F and 3,4-hexanedione in acetonitrile. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

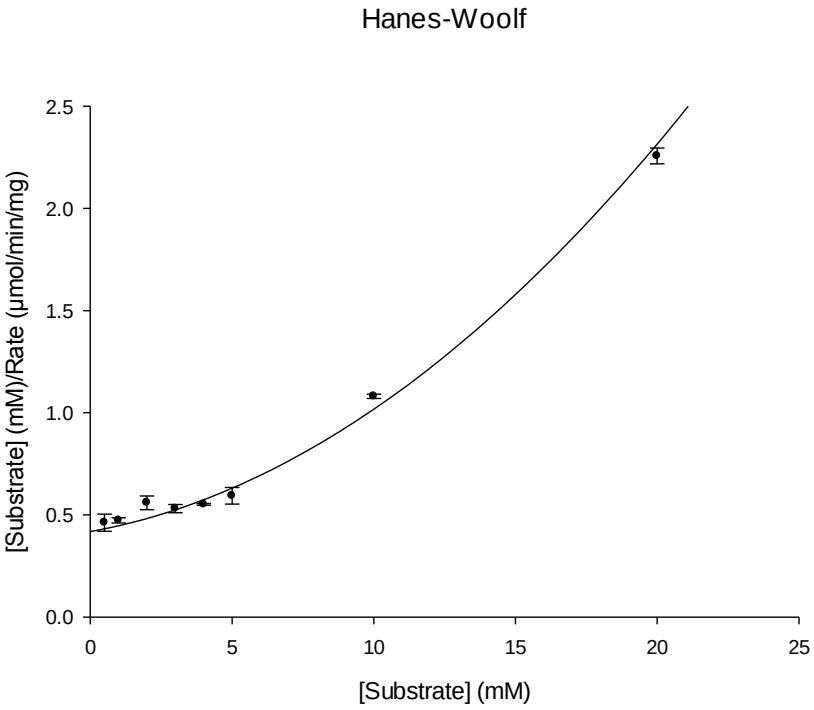


Figure 109 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 3,4-hexanedione in acetonitrile, data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

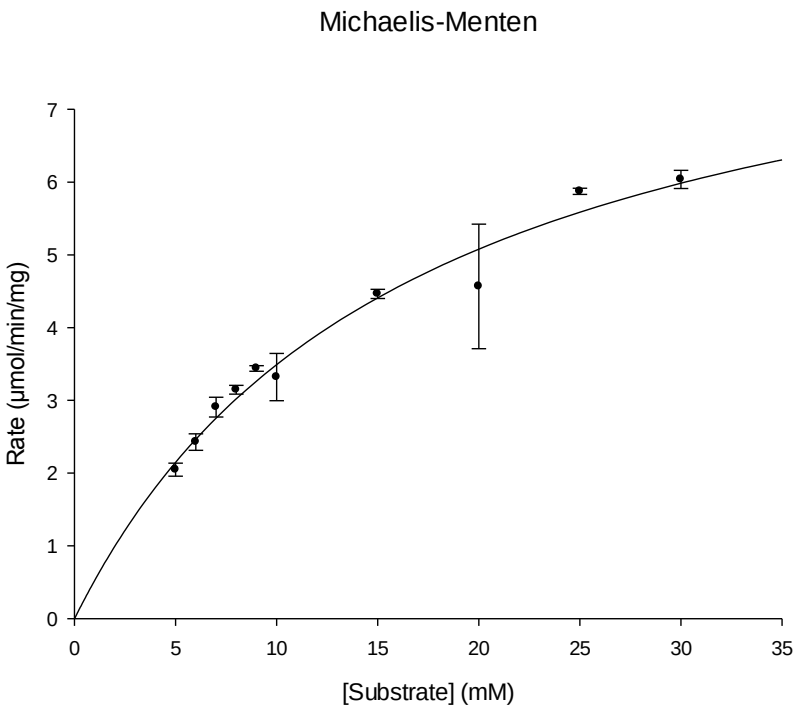


Figure 110 - Graph of rate of reaction vs [S] using ADH F and 3,4-hexanedione in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

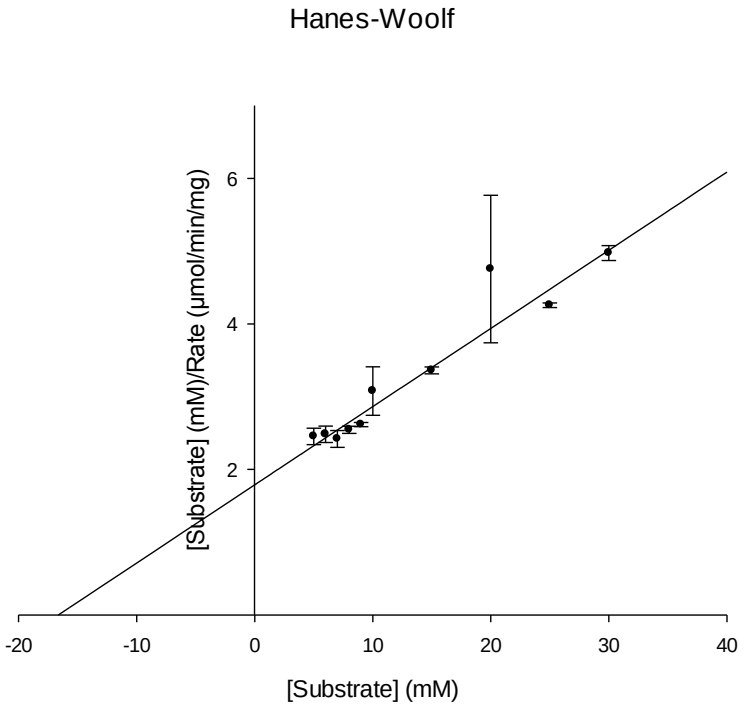


Figure 111 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 3,4-hexanedione in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

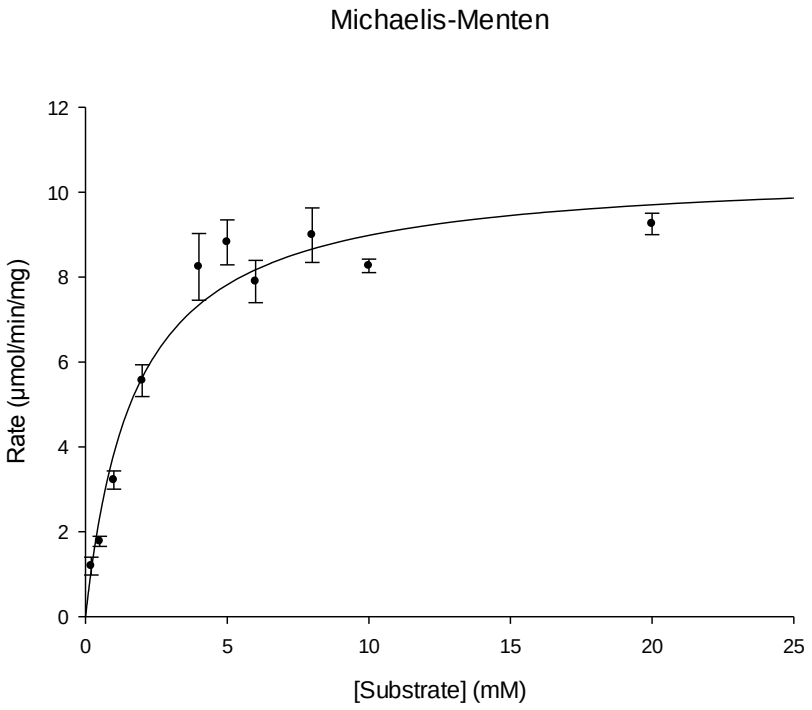


Figure 112 - Graph of rate of reaction vs [S] using ADH F and 2,2,2-trifluoroacetophenone in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

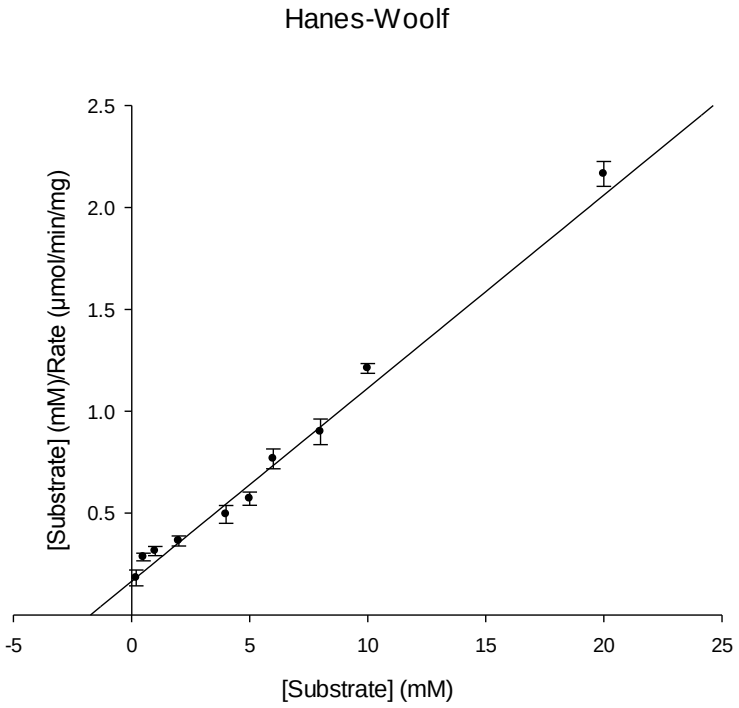


Figure 113 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 2,2,2-trifluoroacetophenone in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

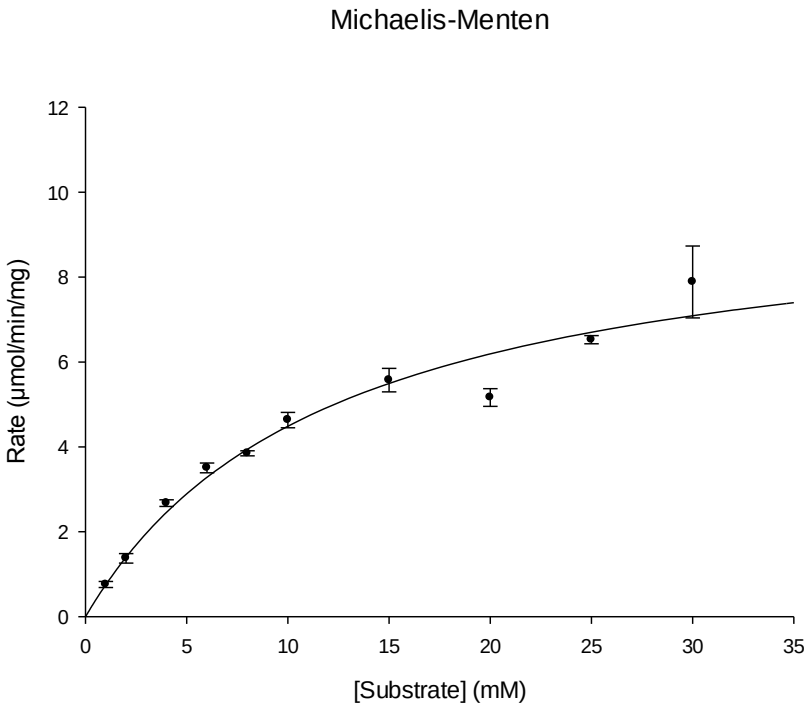


Figure 114 - Graph of rate of reaction vs [S] using ADH F and 2,2,2-trifluoroacetophenone in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

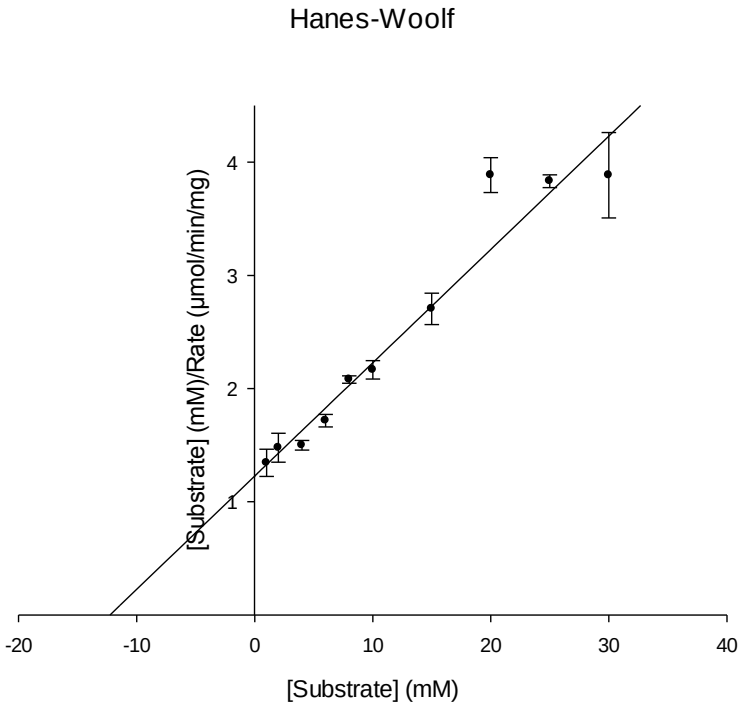


Figure 115 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and 2,2,2-trifluoroacetophenone in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

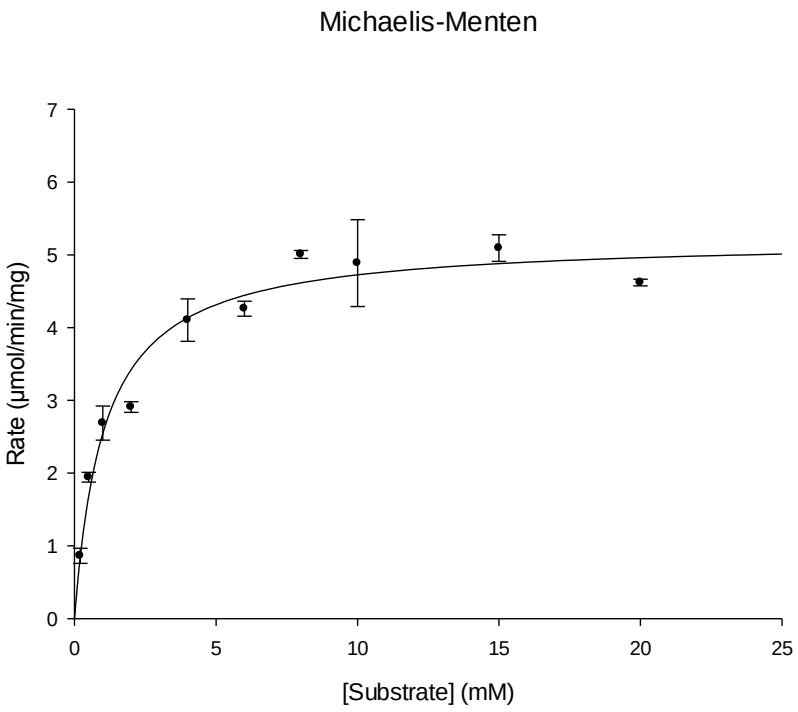


Figure 116 - Graph of rate of reaction vs [S] using ADH F and ethyl 2-oxo-4-phenylbutyrate in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

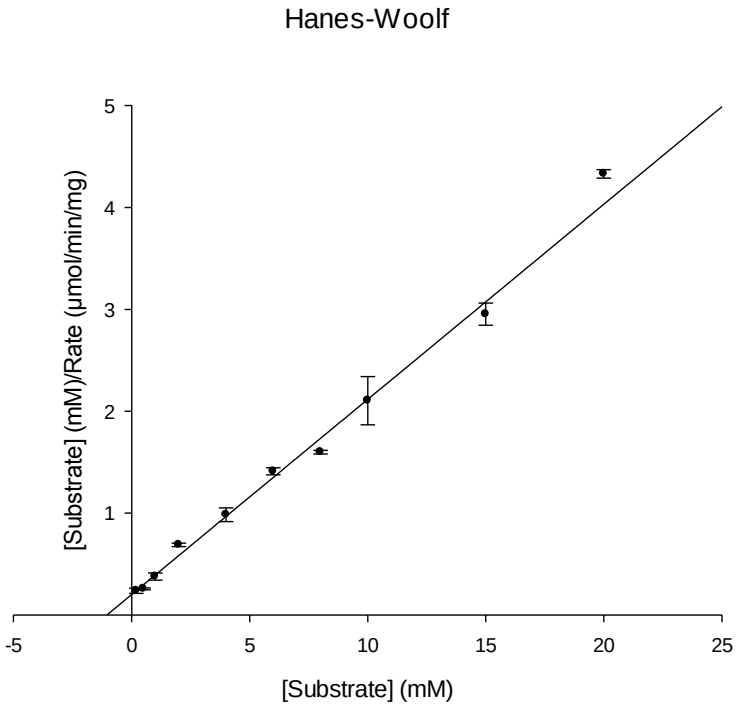


Figure 117 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and ethyl 2-oxo-4-phenylbutyrate in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

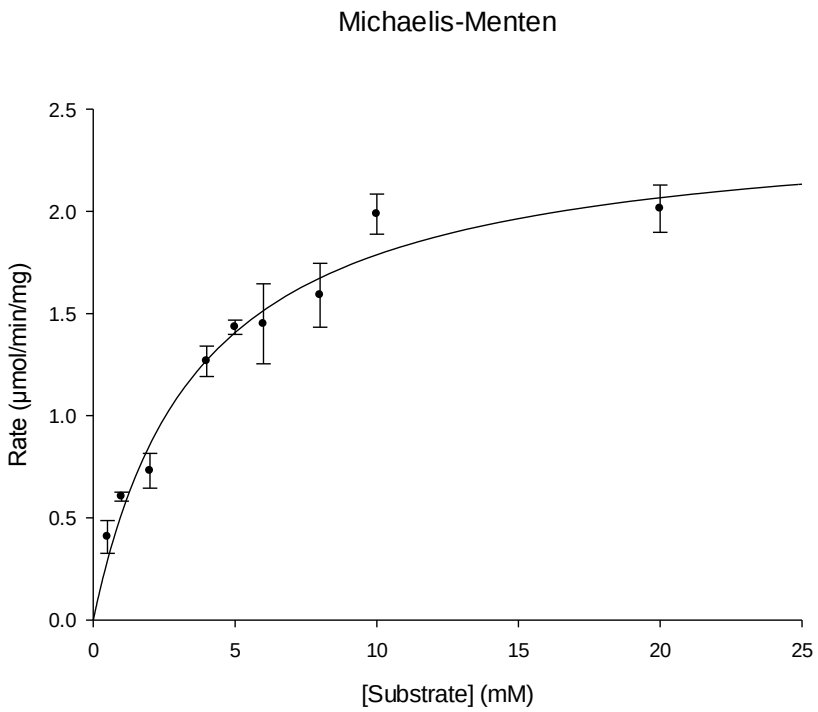


Figure 118 - Graph of rate of reaction vs [S] using ADH F and ethyl 2-oxo-4-phenylbutyrate in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

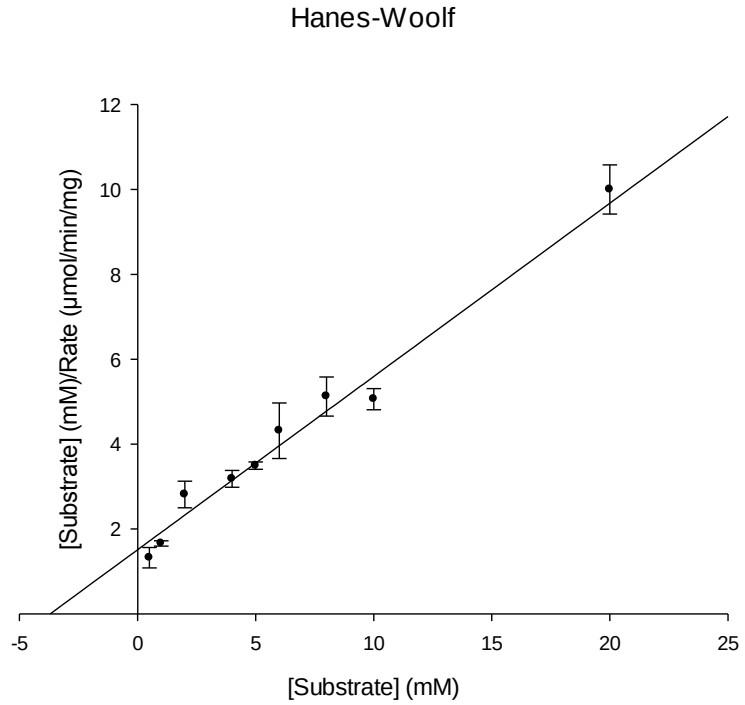


Figure 119 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and ethyl 2-oxo-4-phenylbutyrate in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADH, 50 °C, 1ml assay volume.

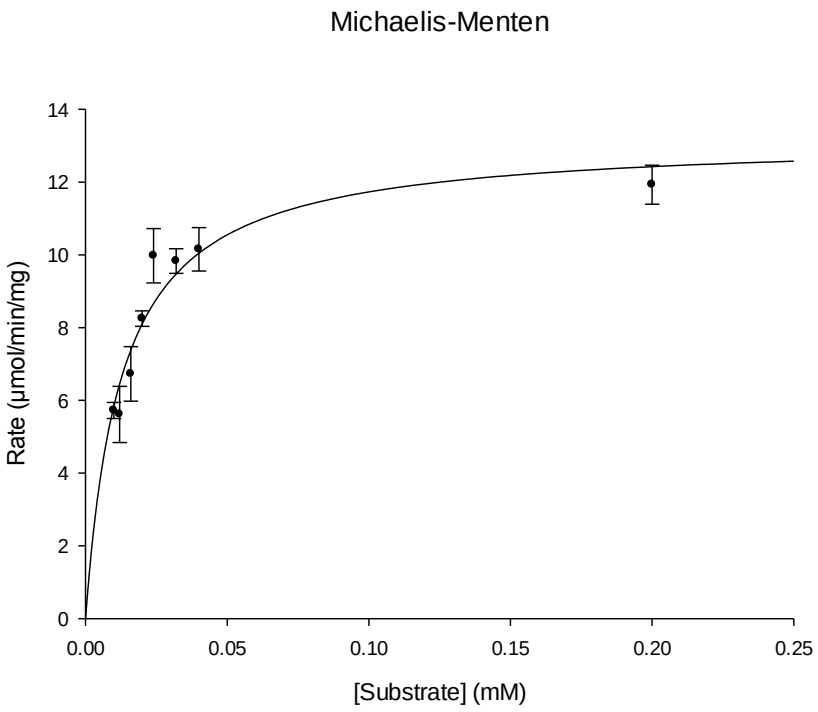


Figure 120 - Graph of rate of reaction vs [S] using ADH F and NADPH. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM 1-phenyl-1,2-propanedione in acetonitrile, 50 °C, 1ml assay volume.

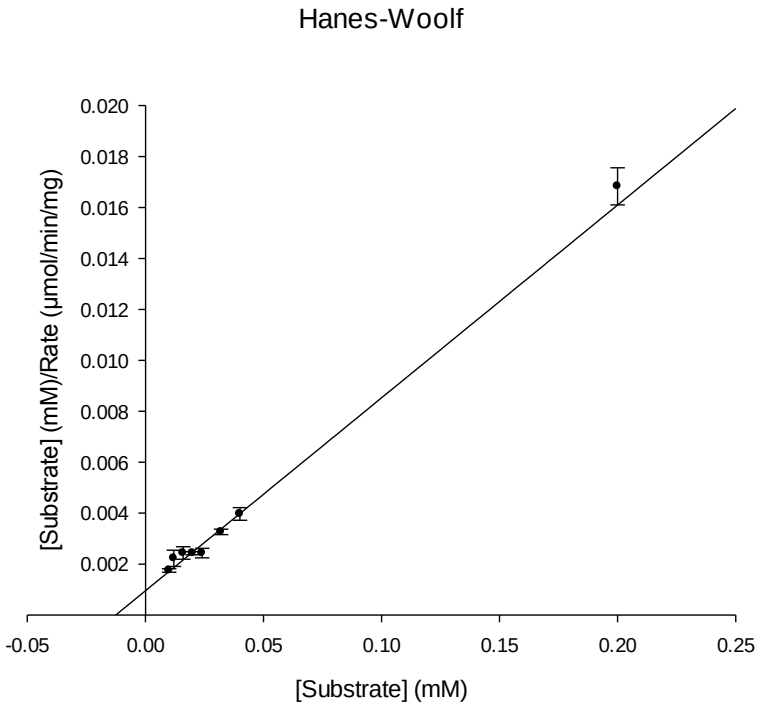


Figure 121 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and NADPH. Conditions: 50mM sodium phosphate pH 7.0, 10mM 1-phenyl-1,2-propanedione in acetonitrile, 50 °C, 1ml assay volume.

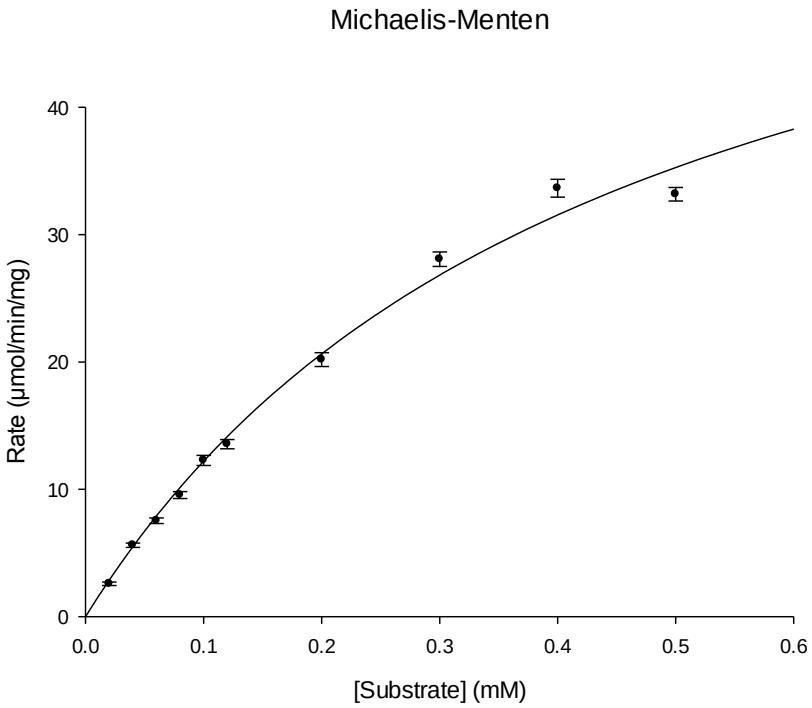


Figure 122 - Graph of rate of reaction vs [S] using ADH F and NADH. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM 1-phenyl-1,2-propanedione in acetonitrile, 50 °C, 1ml assay volume.

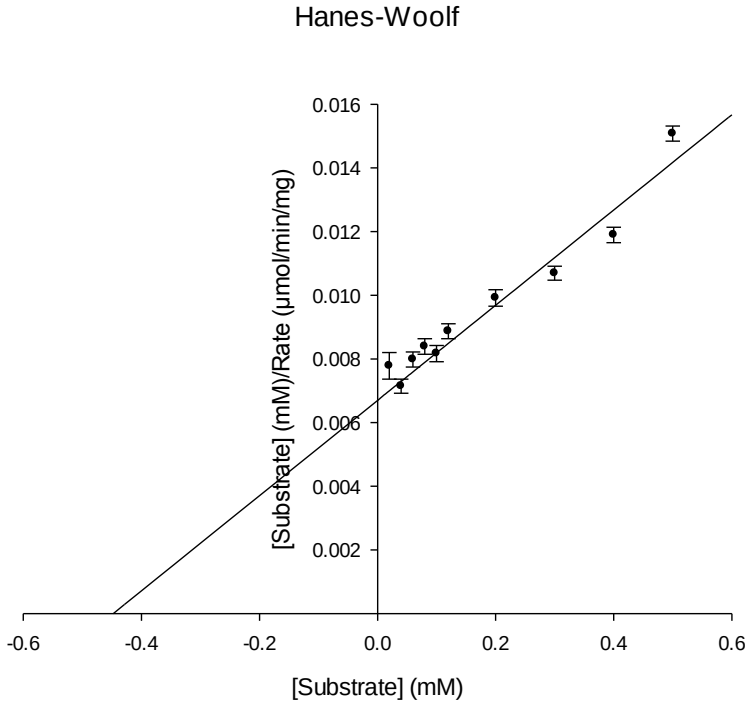


Figure 123 - Hanes-Woolf plot ([S]/v vs [S]) for ADH F and NADH. Conditions: 50mM sodium phosphate pH 7.0, 10mM 1-phenyl-1,2-propanedine in acetonitrile, 50 °C, 1ml assay volume.

Substrate	Cofactor	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
1-phenyl-1,2-propanedione	NADH	28.26±0.94	0.099±0.013	46.8±9.8	960	9700
1-phenyl-1,2-propanedione	NADPH	16.02±0.52	0.010±0.003	18.0±4.3	550	54000
Furfural	NADH	6.58±0.39	10.07±1.61	-	220	22
Furfural	NADPH	10.34±0.37	1.54±0.24	-	350	230
3,4-Hexanedione	NADH	9.31±0.82	16.65±2.83	-	320	19
3,4-Hexanedione[†]	NADPH	40.19±16.90	16.81±8.40	7.11±4.08	1400	81
2,2,2-trifluoroacetophenone	NADH	9.99±0.80	12.27±2.21	-	340	28
2,2,2-trifluoroacetophenone	NADPH	10.55±0.46	1.74±0.29	-	360	206
Ethyl-2-oxo-4-phenylbutyrate	NADH	2.45±0.17	3.70±0.68	-	83	23
Ethyl 2-oxo-4-phenylbutyrate	NADPH	5.22±0.16	1.05±0.16	-	180	169
NADH	-	66.87±3.89	0.44±0.04	-	2300	5200
NADPH	-	13.21±0.70	0.013±0.002	-	450	35000

Table 25 - Summary of kinetic data with ADH F. [†] Results illustrative only. Due to limitation of solvent resistance, it was not possible to get sufficiently concentrated substrate for improved data.

An assortment of noteworthy results was revealed in this area. Initially, the discrepancy between cofactors was exacerbated as first one was preferred, then the other, depending on the substrate involved. Regrettably the cofactor kinetics were done last, which disclosed the root cause – an exceptionally high K_M value for NADH, as well as a much higher V_{max}. Typically cofactors are used at a concentration of 0.2mM, and this work was no exception. This is to minimise absorption, or at least to constrain it to a region where there is highest confidence in the values obtained. By the nature of measuring absorbance figures, particularly the logarithmic scale by which they are measured, it is recommended to remain below an absorbance of 2, and preferentially below 1.5, in order to have absolute confidence in the figure that is recorded. This is compounded when, as is frequently the case during this investigation, substrates also absorb at the monitored wavelength of 340nm. This has led to a generally accepted concentration of 0.2mM NAD(P)H.

Due to frequent substrate absorbance, it was not typical to run any assay with more than 0.2mM cofactor, even if the spectrophotometer was theoretically able to provide accurate readings at that level. Therefore, having a cofactor with a K_M value in excess of 0.4mM was most troubling, as this would countermand the assumption made with all assays of this type that the cofactor is in excess at all times, kinetically speaking. Although it is not ideal to note that the cofactor is not acceptably in excess for these kinetics, it is not an uncommon issue to encounter. Typically a NADH K_M is higher than for NADPH, and it is important to note that ADH F has a much greater affinity for NADPH than NADH. Consequently it isn't surprising given the ability of the enzyme to use both cofactors, that the Michaelis constant for NADH is so high.

The only other issue was where higher substrate concentrations would have been instructive, e.g. 3,4-hexanedione with NADPH as cofactor. The Hanes-Woolf plot indicates substrate inhibition is present; however the kinetics provided by the software package are obviously erroneous and misleading. Forcing a fit to the standard Michaelis-Menten equation without inhibition gives a V_{max} of 11.4 and K_M of 3.0. By visual inspection of the data these kinetic constants are likely to be approximately correct. This was another occasion where an additional reading at 30mM may have been helpful, although further points above this may

have been required to get enough to be analysed appropriately. The reason this was not done was due to the potential effects of the solvent upon the reaction.

It is surprising that Michaelis constants are lower and particularly that specificity constants recorded with 1-phenyl-1,2-propanedione are higher than those recorded for cofactors. Typically cofactor specificity constants are substantially higher than those of the substrates. Whilst native substrates cannot be inferred directly from kinetic data, these data provide a relatively strong signal that whatever the native role of this enzyme, 1-phenyl-1,2-propanedione is a very highly preferred substrate.

3.3.3. STABILITY ANALYSIS

Due to the importance of ADH F, a tertiary screen was conducted, in an identical manner to ADH D, to elucidate potential operating conditions for the enzyme.

Assay Condition	Relative Activity (%)
pH 5.0 Sodium Acetate	47
pH 5.0 Succinate	48
pH 6.0 Citrate	59
pH 6.0 Succinate	47
pH 6.5 MES	71
pH 6.5 Citrate	86
pH 7.0 Sodium Phosphate	100
pH 7.0 MOPS	21
pH 7.5 Sodium Phosphate	82
pH 7.5 Imidazole	80
pH 8.0 Sodium Phosphate	57
pH 8.0 HEPES	48
pH 8.5 Borate	37
pH 8.5 Bicine	29
pH 9.0 Borate	26
pH 9.0 Bicine	23
Unbuffered	77

Table 26 - Tertiary screen with ADH F (buffers) - all tested with 10mM furfural as substrate, NADPH as cofactor, averages of duplicate assays.

First buffers and pH were considered, and in opposition to what was expected, there was no benefit observed of using an acidic pH, at least with the buffers studied. No buffer improved upon the assay buffer, sodium phosphate, with MOPS being a particularly poor buffer, with a five-fold reduction in activity at the same pH as Sodium Phosphate. Whilst there was slight variation depending on the buffer at different pH values, there was almost as big a drop at acidic pH as there was at alkaline, which was quite surprising as generally it would be expected that an acidic pH would increase the activity.

Additive	Dose (vol%)	Relative Activity (%)
10% SDS	1	-
400mM EDTA	0.25	105
	2.5	70
Triton X-100	1	116
	10	101
PEG 600†	1	96
	10	38
DMSO	1	96
	5	74
	10	66
	20	33
	50	-
	93	-
Ethyl Acetate*	1	90
	5	77
	10	52
	20	49
Acetonitrile	1	93
	5	76
	10	58
	20	35
	50	7
	93	-
Cyclopentyl Methyl Ether*	1	88
	5	73
	10	60
	20	64
	50	(Solvent Distortion)
Hexane*	1	116 (115)
	5	115 (110)
	10	115 (104)
	20	125 (100)
	50	155 (78)
	93	(Solvent Distortion)

Table 27 - Tertiary screen with ADH F (additives and solvents) - all tested using 10mM furfural as substrate, NADPH as cofactor and pH 7.0 sodium phosphate. *Ethyl acetate, CPME and hexane all form bilayers, impacting upon assays. For example, ethyl acetate above 8% forms a bilayer, indicating solvent saturation, potentially affecting observed rates and at high levels distorting assays. Normalised activities are provided in brackets. †PEG 600 control registered a rate of 16%. All values averages of duplicate assays.

Additives had a negative effect on rates of activity, with SDS again suppressing all activity. EDTA was tolerated at low concentration, but at the higher concentration tested reduced activity by 30%. Triton was well tolerated, but PEG600 notably reduced activity at the higher concentration tested.

The solvent study is of particular interest, given that these were not the first assays conducted with ADH F. In this case, with furfural as substrate, all solvents aside from hexane are poorly tolerated. Hexane may even provide a boost at lower quantities, with a 15% increase in activity noted at 1vol%. The higher readings for increased volumes are likely an artefact of the immiscibility of the solvent in the buffer. The reason furfural was chosen for this study was

because of its solubility in the buffer, hence allowing all the assays to be conducted from the same stock solution.

The problem is that this work does not compare to earlier solvent work which was done in order to carry out previous work that required substrates to be dissolved in solvent. The first initial solvent screen was carried out using butanal as substrate, again due to ease of use. This work indicated no particular solvent resistance for DMSO or ethyl acetate; however, those assays were done using PMMA cuvettes, which were determined to lack solvent resistance themselves.

Upon repetition of the work in quartz cuvettes in preparation for the secondary screen, DMSO appears to have little effect at a concentration of 10vol%, ethyl acetate at 1%, and acetonitrile at 1%. On that basis, the secondary screen uses DMSO to produce stock solutions as required. Upon moving to kinetics, it was immediately apparent that DMSO does inhibit, when using 1-phenyl-1,2-propanedione as substrate.

More rigorous testing with that substrate reveals that DMSO is only suitable for kinetics analysis, i.e. no observed effect on the activity, up to 1vol%, and in fact acetonitrile is the most suitable solvent with no observed effect up to 2vol%.

Comparing the latter experiments with the entries in Table 27 it would seem that there is more variation than would be expected. Whilst the focus is different between an operational solvent screen whereby small amounts of solvent are added and no effect is permitted, and a high level broad scan of solvent ratios, there shouldn't be a huge variance between the two where they overlap. As the difference between the two is in the substrate choice, this could well be the important discrepancy.

Furfural and butanal are notably weaker substrates for ADH F than 1-phenyl-1,2-propanedione; therefore, the potential implication is that apparent solvent resistance is dependent on the relative strength of the substrate. Stated more accurately, for a given enzyme, stronger substrates may provide protection against solvent denaturation through by being more likely to be bound to the enzyme at any given point. More frequent and stronger binding may strengthen the enzyme against denaturing effects. This would be apparent as substrate dependent solvent effects, and is something that should be investigated further.

In terms of overall solvent resistance, with all solvents tested there was a 50% residual activity after adding either 10vol% or 20vol%, which indicates potential usefulness with these solvents as co-solvents. Of particular interest would be CPME, as this solvent is absolutely immiscible with water, marketed as such and as the next generation green solvent. Therefore, of all the substrates tested, having substantial residual activity in this solvent may be the best advantage.

Ion	Dose (vol%)	Relative Activity (%)
100mM $FeCl_2$	1	92
	10	51
100mM $ZnCl_2$	1	107
	10	117
100mM $MnCl_2$	1	106
	10	102
100mM $MgCl_2$	1	104
	10	108
100mM $NiCl_2$	1	106
	10	102

Table 28 - Tertiary screen with ADH F (ions) - All tested using 10mM furfural and pH 6.5 citrate buffer. Necessary to alter buffer composition due to reaction of ions with sodium phosphate buffer. Note high absorbance of ions at 10mM concentration.

As noted with ADH D, the choice of buffer was distinctly unhelpful when considering ions; however, it appears that only iron and zinc had an appreciable effect upon activity. Zinc with a minor increase in activity and iron with a substantial decrease would indicate that ADH F is zinc dependent. This correlates well with EDTA having a detrimental effect upon activity.

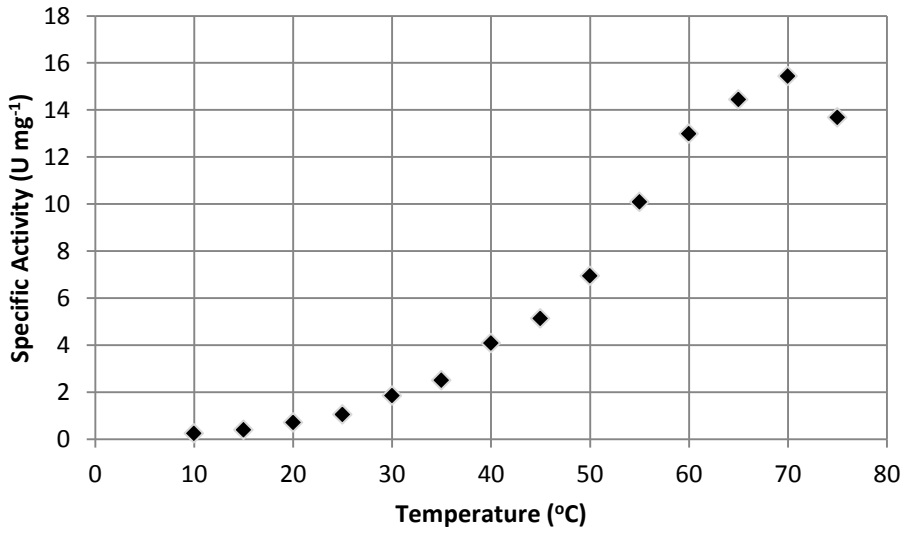


Figure 124 - Optimal temperature study with ADH F using NADPH as cofactor and butanal as substrate

The optimal temperature study indicated an optimal temperature of 70 °C. Compared to the operating temperature of *G. thermoglucosidasius*, this would mean that this enzyme could work well across the entire range, with only a minor loss at 75 °C, and broadly the same activity at 60 °C as at 75 °C.

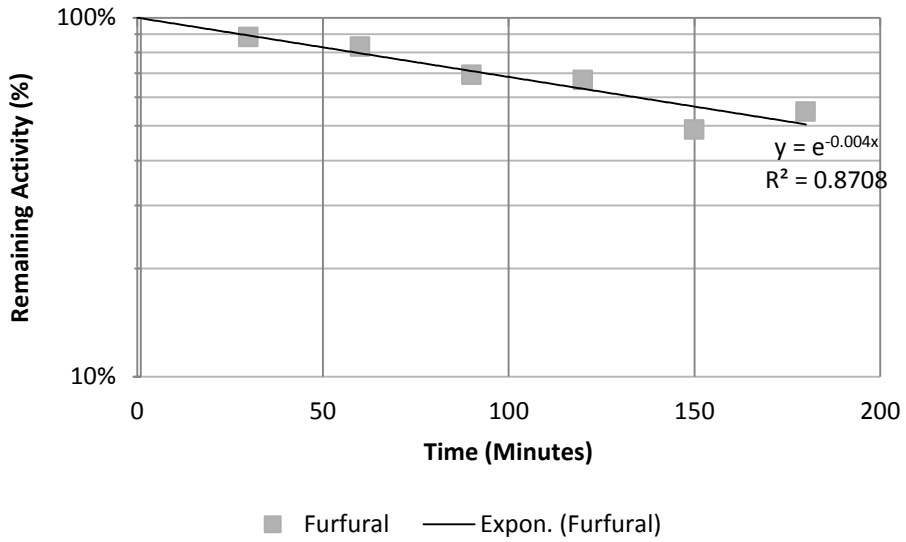


Figure 125 - Preliminary thermostability test, conducted at 60 °C, with 10mM furfural as substrate and NADPH as cofactor.

Initial thermostability experiments were conducted over three hours; in contrast to previous studies carried out at 70 °C or higher, this was carried out at 60 °C, the standard growth temperature of *G. thermoglucosidasius* as preliminary work had determined low stability at higher temperatures.

$$t_{1/2} = \frac{ln(2)}{0.004}$$

$$t_{1/2} = 173\ mins$$

With an estimated half-life of approximately three hours, long term stability at this temperature would appear to be limited. Compared to ADH D, there is a substantial reduction in stability, from a half-life of 38.5 hours at 70 °C to three hours at 60 °C.

On the basis of this data an extended thermostability experiment was carried out, due to the potential of this enzyme. In addition glycerol was added to a final quantity of 50vol% with a duplicate test. As glycerol is known to improve the stability of enzymes in some cases, this was done to indicate whether the addition of additives may augment the thermostability of the enzyme. Furthermore, by adding the glycerol it was anticipated that this may aid in reducing intrapoint variability by increasing the viscosity and therefore dispersing the enzyme more fully within the aliquot.

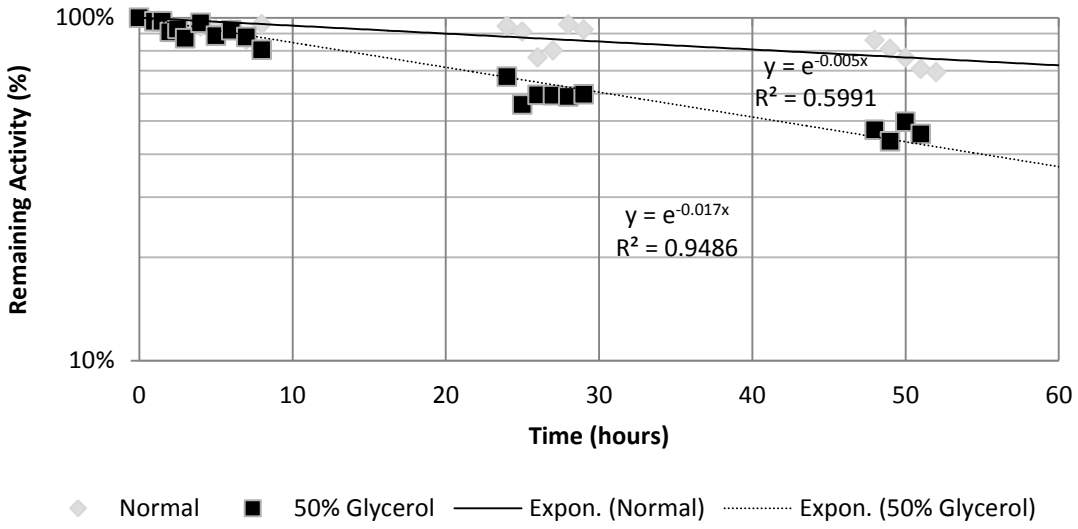


Figure 126 - Extended thermostability test at 60 °C using furfural, NADPH as cofactor.

Over the course of 52 hours aliquots were tested for activity, with neither the standard or 50vol% glycerol samples degrading at the same level as noted previously. Although there was reduced variation in the glycerol run, the run without glycerol was substantially more stable across the time period analysed. Even with the heightened variation, it is clear that ADH F is more stable without additives.

Using both figures to provide estimates of stability:

$$t_{1/2} = \frac{ln(2)}{0.017} = 41\ hours$$

$$t_{1/2} = \frac{ln(2)}{0.005} = 139\ hours$$

Thus the estimate for ADH F stability at 60 °C is 41 hours when 50vol% glycerol is added, and 139 hours when not. This is much more in line with ADH D, and represents substantial stability at the likely native temperature.

The thermal shift assay for ADH F indicated an inflection point at 76.0 °C, slightly above the previously determined T_{opt} of 70 °C.

3.3.4. CHEMICAL ANALYSIS

Finally, assays with ADH F were analysed by GC/MS as previously, to determine conversions produced during these assays and to provide an indication of the products produced. Due to the previously uncovered activity with multiple substrates of interest, this is a particularly important area of study.

	Conversion			
Substrate	**1 Hour**		**16 Hours**	
	NADH	**NADPH**	**NADH**	**NADPH**
5-Norbornene-2-carboxaldehyde*	13%	78%	18%	36%
Butanal**	7%	22%	5%	13%
Furfural	30%	44%	Activity	Activity
Methyl benzoylformate	-	-	-	-
Decanal	0%	1%	0%	0%
Benzaldehyde	0%	3%	Activity	Activity
4-Chlorobenzaldehyde	5%	17%	8%	35%
2,3-Pentanedione	N/A	N/A	5%/23%	9%/45%
2,3-Butanedione	N/A	N/A	25%	58%
Ethyl-2-oxo-4-phenylbutyrate	33%	39%	62%	64%
4-Phenyl-2-butanone	0%	0%	0%	0%
1-Phenyl-1,2-propanedione	9%	16%	18%/6%	N/A
2',3',4',5',6'-Pentafluoroacetophenone	2%	7%	3%	14%
Acetophenone	N/A	N/A	1%	1%
Ethyl 4-chloroacetoacetate	-	-	-	-
Ethyl acetoacetate	N/A	N/A	N/A	6%
2,2,2-trifluoroacetophenone	N/A	N/A	14%	48%

Table 29 - Summary of GC/MS data with ADH F. N/A indicates an assay was not run, "-" indicates substrate unsuitable for GC/MS analysis. *1 hour assay was run using double enzyme, ** substrate evaporation suspected.

Allowing for the same problems with analysing data as previously noted, with respect to furfural, decanal and benzaldehyde, activity was confirmed with all substrates with the exception of 4-phenyl-2-butanone, which was expected. Interestingly, despite a lack of activity noted in previous assays, activity was noted with acetophenone.

In keeping with previous work, all assays were conducted with both cofactors independently. The only exceptions to this were 1-phenyl-1,2-propanedione with NADPH and ethyl acetoacetate with NADH, in these instances the data were unrecoverable. It is also important to note that 0.2mM cofactor was used, which appears to bias conversion in favour of NADPH, with the sole exception of ethyl 2-oxo-4-phenylbutyrate where there is approximately the same conversion after one hour and 16 hours. This could only occur if the limitation is somehow cofactor based, i.e. if cofactor regeneration was not a contributing factor, then it would be expected that all assays would have roughly the same conversion irrespective of cofactor used.

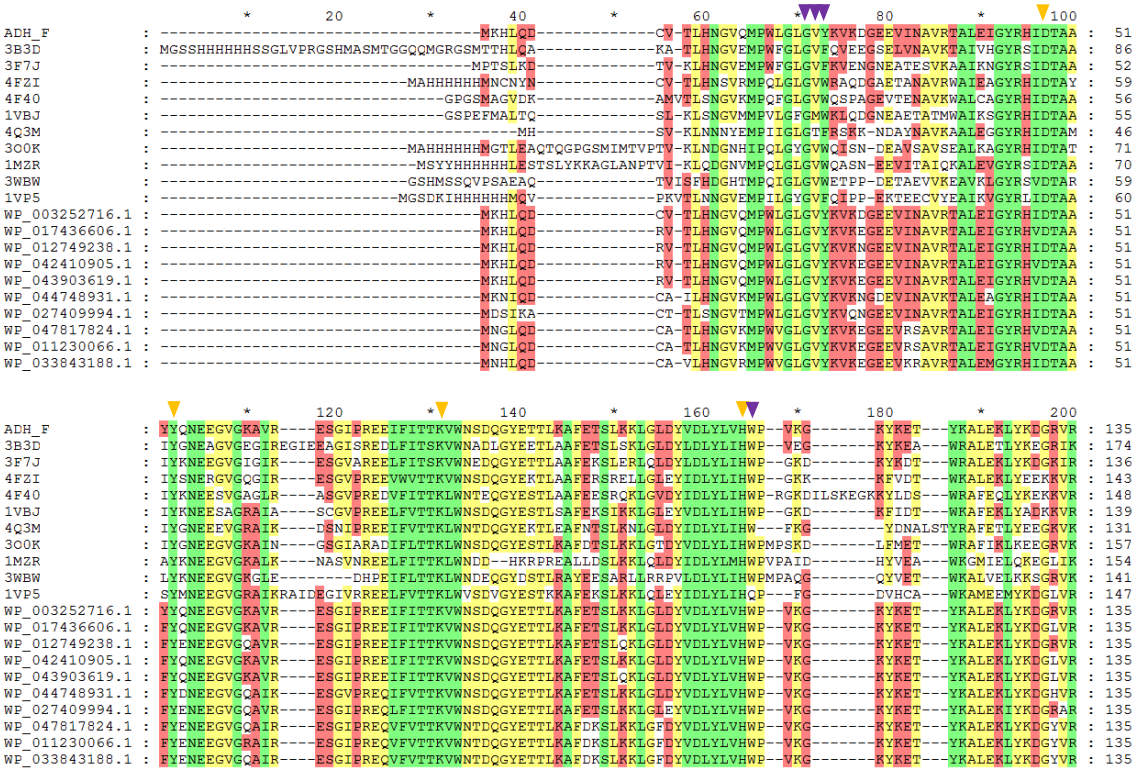
Proceeding on the basis that both ADH F and glucose dehydrogenase enzymes used in the assays are thermostable, then that leaves either the substrates or the cofactor itself as not thermostable. At 50 ^{o}C glucose is stable and the majority of substrates tested are also stable, the prominent exception being butanal. Accordingly, as previously suggested, it would appear that cofactor stability is the major issue which is not allowing complete conversion. All the other aspects may degrade partially over time, particularly over 16 hours, but it appears implausible they would prohibit complete conversion; the only other issue would be the position of equilibrium.

Similar to ADH D, there was a 3:1 ratio in observed product with 1-phenyl-1,2-propanedione, with the larger proportion being 1-hydroxy-1-phenyl-2-propanone. No diol was observed. Contrastingly though, when 2,3-pentanedione was considered, instead of a 4:1 ratio as per ADH D, ADH F displayed a 5:1 bias in favour of 2-hydroxy-3-pentanone. Again, no diol was observed. Lastly, with regards to 2,3-butanedione, only acetoin was noted, no diol.

The lack of activity with both ketone groups on the same molecule, even where diketones are preferred to the exclusion of monoketones, cannot be a result of cofactor degradation. Even where cofactor is limiting, there would still be a small amount of diol present. Whilst not impossible, it is exceptionally unlikely that the diol could be present, but below the limit of detection on the grounds that alcohols were virtually always detected with signal strength an order of magnitude greater than their equivalent ketones.

3.3.5. SEQUENCE ALIGNMENT

After ADH E, this sequence alignment is much more in keeping with the previous alignments, with a high level of sequence similarity, particularly with the most similar proteins in the NCBI RefSeq Database. Particular highlights though are the two most similar crystal structures, which have >60% identity with ADH F due to both deriving from *B. subtilis*.



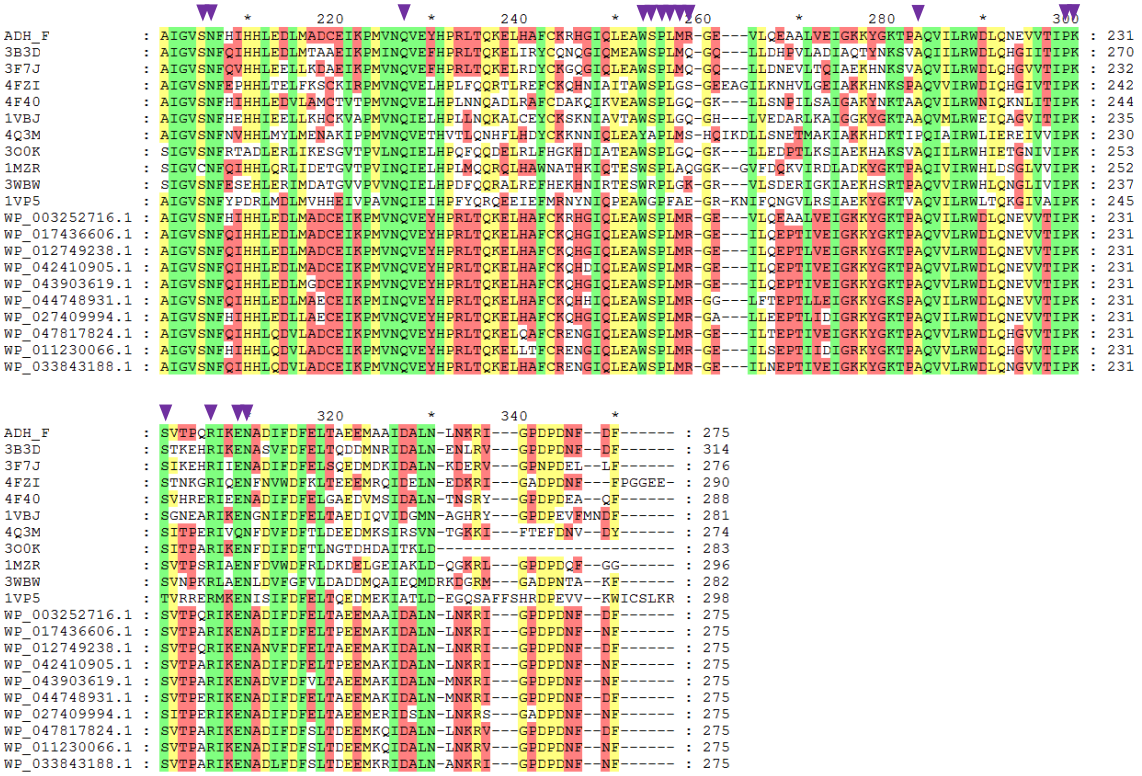


Figure 127 - Multiple Sequence Alignment of ADH F with most similar protein sequences associated with crystal structures in the PDB and most similar protein sequences from the NCBI RefSeq Database obtained using the NCBI Seqr Sequence Search. Green shading indicates 100% conservation, yellow shading indicates >80% conservation and red shading indicates >50% conservation of amino acids across selected sequences. Labels refer to crystal structures from the PDB or proteins from the NCBI RefSeq Database. All similar proteins are annotated as glyoxal reductases from the genera *Geobacillus, Bacillus,* and *Anoxybacillus.* Crystal structures are annotated as: 3B3D is an AKR (YtbE) from *B. subtilis*, 3F7J is an AKR (YvgN) from *B. subtilis*, 4FZI is a prostaglandin F synthase from *Typanosoma cruzi,* 4F40 is a prostaglandin F synthase from *Leishmania major*, 1VBJ is a prostaglandin F synthase from *Trypanosoma brucei*, 4Q3M is an AKR from a Medee basin deep-sea metagenome, 3O0K is an AKR from *Brucella melitensis,* 1MZR is an AKR (DkgA) from *E. coli*, 3WBW is an unknown oxidoreductase (Gox444) from *Gluconobacter oxydans*, 1VP5 is a 2,5-diketo-D-gluconic acid reductase (TM1009) from *Thermotoga maritima*. Purple triangles indicate the active site residues and orange triangles indicate the catalytic triad. All these sites provided via NCBI Conserved Domain Database and obtained through the Conserved Domain Search. Sequences aligned using MEGA 6.0 (Tamura et al., 2013) with MUSCLE and visualised using Genedoc 2.7 (Nicholas et al., 1997).

3B3D and 3F7J are NADPH dependent AKRs from *B. subtilis*, also known as YtbE (AKR5G2), and YvgN (AKR5G1). The former is an essential enzyme, i.e. required for growth, as it is necessary for the removal of toxic aldehydes (Commichau et al., 2013). The latter is involved with sporulation, but both have broad substrate scopes, from glyoxal to benzaldehyde derivatives. They have high (~70%) sequence identity with each other, in addition to being the most similar crystal structures to ADH F (Lei et al., 2009).

As has been mentioned previously, many AKRs have a “safety belt” to secure the cofactor. In such enzymes, the conformational change of the safety belt to release the $NADP^+$ is believed to be rate-limiting. In YvgN the residues involved are K26 and Q194, corresponding to K74 and Q258; no significant conformational change was noted. In the case of YtbE, as there was a different residue (K74Q), no belt was observed (Lei et al., 2009). ADH F has K74, but Q258R, and therefore isn’t clear *a priori* whether ADH F has a safety belt. The catalytic triad is as annotated in Figure 127: D97, Y102, K131 and H164. Unsurprisingly, these residues are identical throughout the MSA.

Although both YvgN and YtbE have a breadth of substrates, they prefer ortho-substituted rather than para-substituted benzaldehydes. It is believed that N134, which is conserved in ADH F, may interact with the 4-group and thereby influence substrate specificity. Of the substrates examined both enzymes preferred 2-chlorobenzaldehyde and 2,4-chlorobenzaldehyde and less preferred benzaldehyde, methylglyoxal and glyceraldehyde as substrates. No activity at all was detected with sugars such as glucose and fructose (Lei et al., 2009).

YtbE in particular has been taken forward for additional investigation, extending its substrate scope to include 2-chloroacetophenone, 2,2,2-trifluoroacetophenone, ethyl 4-chloroacetoacetate and ethyl pyruvate, among others. It was also determined to be monomeric in structure. Another important feature is that the enzyme is exceptionally enantioselective, affording several substrates tested (including ethyl acetoacetate, ethyl pyruvate) as >99% enantiomeric excess. With the exception of ethyl-2-oxo-4-phenylbutyrate, all products are *S*-alcohols (Ni et al., 2011). Latterly YtbE has been investigated to perform the asymmetric reduction of *o*-chlorobenzoylformate to *(R)*-*o*-chloromandelate as a key intermediate in the production of *(S)*-clopidogrel, a medication for atherosclerosis. In particular stabilising additives have been sought, to reduce the rate of degradation in preparation for an industrial scale operation. Another route considered was to drop the operating temperature as low as 20 °C to improve the lifetime of the enzyme (Xu et al., 2014). Given the demand for this process and the poor thermostability of YtbE it would seem to be a high priority route for future investigation. With the enhanced thermostability of ADH F and the demonstrated activity with similar substrates, this is a path of interest.

4FZI, 4F40 and 1VBJ are prostaglandin F synthases, another type of NADPH dependent AKR. 4FZI is the apo form of a prostaglandin F synthase from *Trypanosoma cruzi*, 4F40 is the apo form of a prostaglandin F synthase from *Leishmania major*, and 1VBJ is the same enzyme from *Trypanosoma brucei*. Position 258 is key to stabilising the phosphate of NADPH, whether a serine or glutamine residue, which is important due to the relative flexibility of this area, which is exacerbated by a conformational change upon cofactor binding. Thus by virtue of the cofactor binding to the enzyme becomes more stable (Moen et al., 2015). It is worth noting that this is the same residue referred to in YtbE and YvgN studies as being relevant to the "safety belt".

To conclude, ADH F is the only ADH studied that has such similar crystal structures, due mainly to their being from *B. subtilis*. With such closely related structures it would be prudent to use them as an example, particularly as the major noted problem with YtbE was a lack of thermostability, whereas a thermophilic enzyme is substantially more likely to be suitable. As the catalytic tetrad is well defined and conserved, it is doubtful whether progress can be made in terms of determining any structural features, at least until a crystal structure is obtained. Two residues that may be of interest are M257 and R258, which are the only residues to be annotated as part of the active site and not be conserved, thus potentially those residues may be key to the substrate scope of ADH F.

3.3.6. SUMMARY

ADH F is by far the most interesting alcohol dehydrogenase isolated as part of this investigation. With exceptional substrate scope and relatively high activity across the range, and the ability to use both NADH and NADPH, it is a promising enzyme for industrial use. Activity has specifically been confirmed with aliphatic, cyclic and aromatic aldehydes, diketones, alpha and beta ketoesters, as well as chlorinated and fluorinated examples of the above.

Further investigation is required in order to determine the exact products produced, and engineering of the enzyme may be necessary to promote a preferred product over another where multiple ketones are present on a substrate. As monoketones are not substrates for this enzyme, in the absence of another electron withdrawing group, it would seem to be obligatory for that group to be present to enable reaction. Thus, whilst it may prove that diketones, in particular 1-phenyl-1,2-propanedione, may be the preferred substrates for this enzyme, it may be beneficial to utilise this enzyme with substrates that only have one available ketone to reduce by-product formation.

Some activity loss with higher levels of solvent has been noted, but this is less relevant when high levels of conversion are currently stymied by cofactor inactivation rather than enzyme degradation.

3.4. ADH G

The final ADH studied in detail is ADH G. Although expressed well, it did not purify cleanly, as shown below. The band corresponding to the larger protein was believed to be the actual protein; however, after considering the sequence alignments in Section 3.4.5, it can be seen that the original sequence was inaccurate. Specifically the start codon was misannotated, leading to the incorporation of 34 additional amino acids at the N-terminus of the protein, which likely includes a ribosome binding site, and leading to two produced proteins. Thus the *smaller* band indicates the expected protein, and the larger includes these extra amino acids. For the purposes of activity, it was assumed that all protein was identical and active, an assumption that is not ideal, but suffices for an initial investigation.

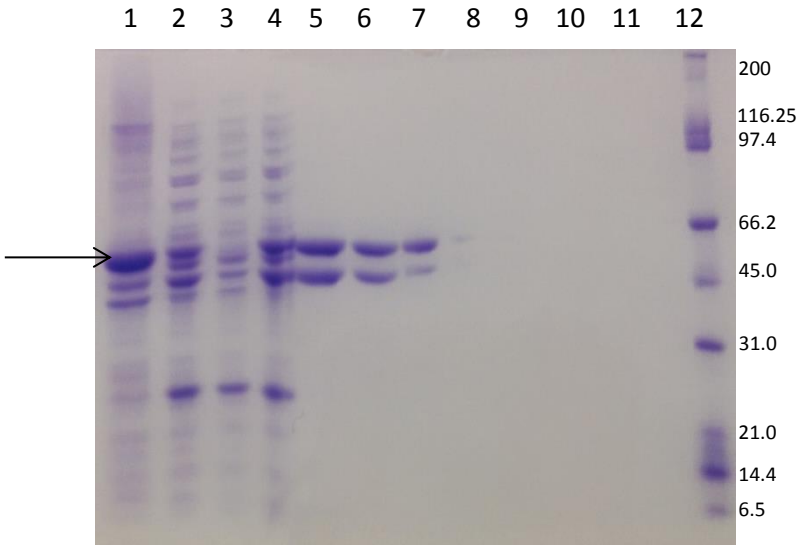


Figure 128 - SDS PAGE gel of ADH G. Lane 1: insoluble fraction, lane 2: soluble fraction, lane 3: 2nd flowthrough, lane 4: 0% His-Elute Buffer, lane 5: 1%, lane 6: 2.5%, lane 7: 5%, lane 8: 10%, lane 9: 30%, lane 10: 100% His-Elute Buffer, lane 12: protein markers. The numbers on the right indicate relative molecular masses (M_r) of markers in kDa; the arrow indicates the putative ADH, with an expected M_r of 49.4kDa.

3.4.1. SPECIFIC ACTIVITIES, SUBSTRATE SCOPE, CONVERSIONS

As is usual, a primary and secondary substrate screen was run, with the primary substrate screen using the same substrates as previous.

Based upon preliminary work it was determined that ADH G utilised NADPH exclusively, and therefore this cofactor was used throughout. From the primary screen it was determined that the enzyme was an aldehyde reductase, with no ability to use ketones as substrates. All aldehydes tested had some activity, although this varied. Therefore, for the secondary screen several linear aldehydes and all other non-linear aldehydes were tested, to determine the scope of this enzyme.

Substrate	Specific Activity ($U\ mg^{-1}$)
Butanal	8.3
Decanal	6.6
2-Butanone	-
2-Decanone	-
3-Pentanone	-
Isobutanal	0.5
Cyclohexanone	-
Furfural	4.6
Acetophenone	-
Ethyl acetoacetate	-
Benzaldehyde	2.3
Pyruvate	-
Ethyl pyruvate	-
Ethyl levulinate	-
Levulinate	-
α-Tetralone	-
3-Methyl-2-butanone	-
2,3-Pentanedione	-
Ethyl 2-oxo-4-phenylbutyrate	-
4-Phenyl-2-butanone	-
1-Phenyl-1,2-propanedione	-
2,2,2-Trifluoroacetophenone	-
Ethyl 4-chloroacetoacetate	-
Ethyl 4,4,4-trifluoroacetoacetate	-
2',3',4',5',6'-Pentafluoroacetophenone	-

Figure 129 - Primary screen of ADH G with NADPH as cofactor.

Substrate	Specific Activity ($U\ mg^{-1}$)
4-Chlorobenzaldehyde	5.1
Methanal	0.6
Propanal	2.3
Hexanal	9.4
Octanal	12.2
4-Nitrobenzaldehyde	(Substrate Interference)
Glyoxal	2.6
5-Norbornene-2-carboxaldehyde	6.5

Figure 130 - Secondary screen with ADH G and NADPH as cofactor.

With the notable exception of 4-nitrobenzaldehyde, which interfered with the assay, all aldehydes tested throughout had activity, including methanal, glyoxal and large aldehydes. Thus, although activity noted was relatively low, rarely above 10 U mg^{-1}, due to the very large substrate range, this is only to be expected.

3.4.2. KINETICS

Three of the more promising aldehyde substrates were taken forward to analyse their kinetics. As the enzyme was not able to use ketone substrates, it was not considered a priority for investigation.

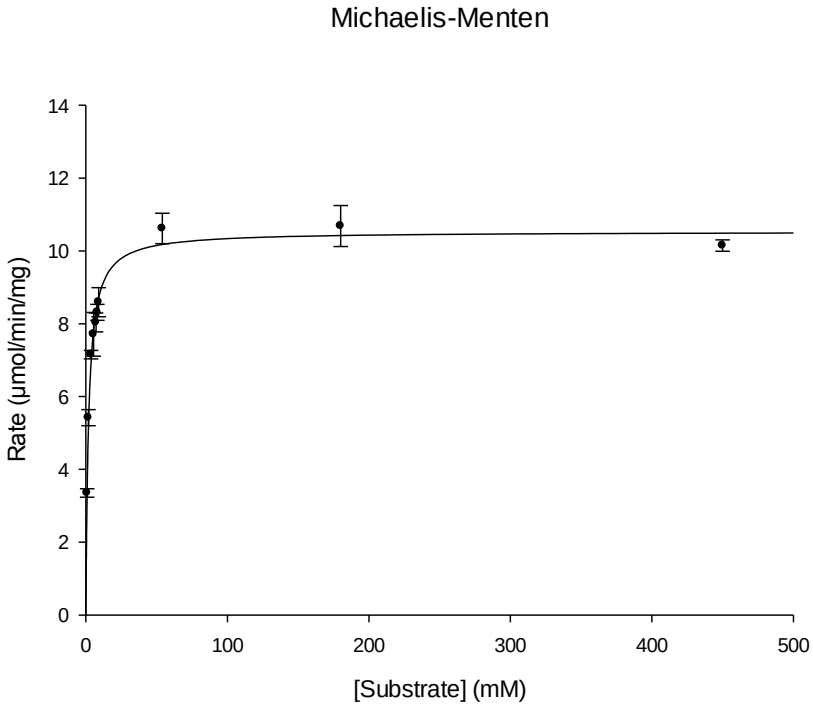


Figure 131 - Graph of rate of reaction vs [S] using ADH G and butanal. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

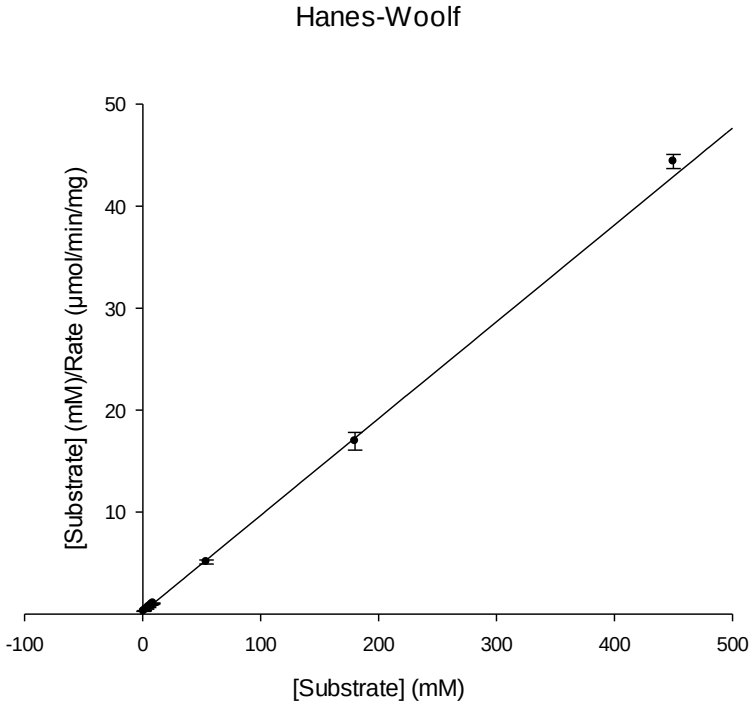


Figure 132 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and butanal. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

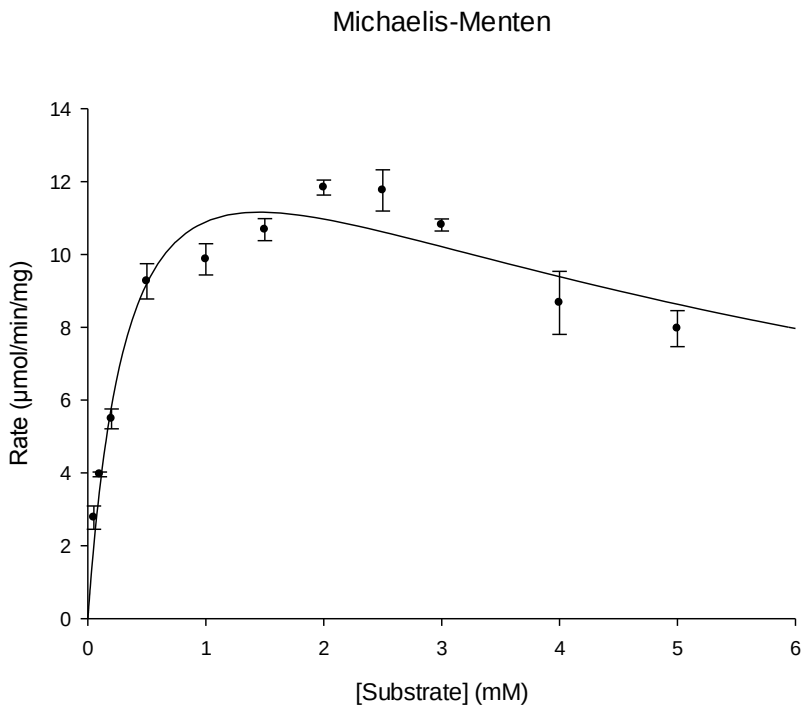


Figure 133 - Graph of rate of reaction vs [S] using ADH G and octanal in acetonitrile. Data fitted to a Michaelis-Menten curve, assuming substrate inhibition, using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

Figure 134 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and octanal, data fitted assuming substrate inhibition. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

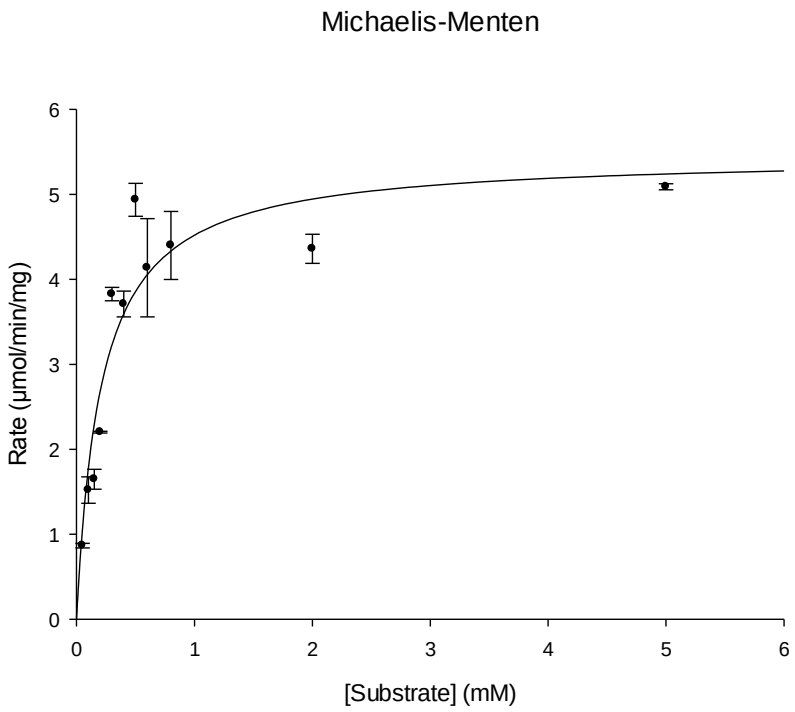


Figure 135 - Graph of rate of reaction vs [S] using ADH G and 5-norbornene-2-carboxaldehyde in acetonitrile. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

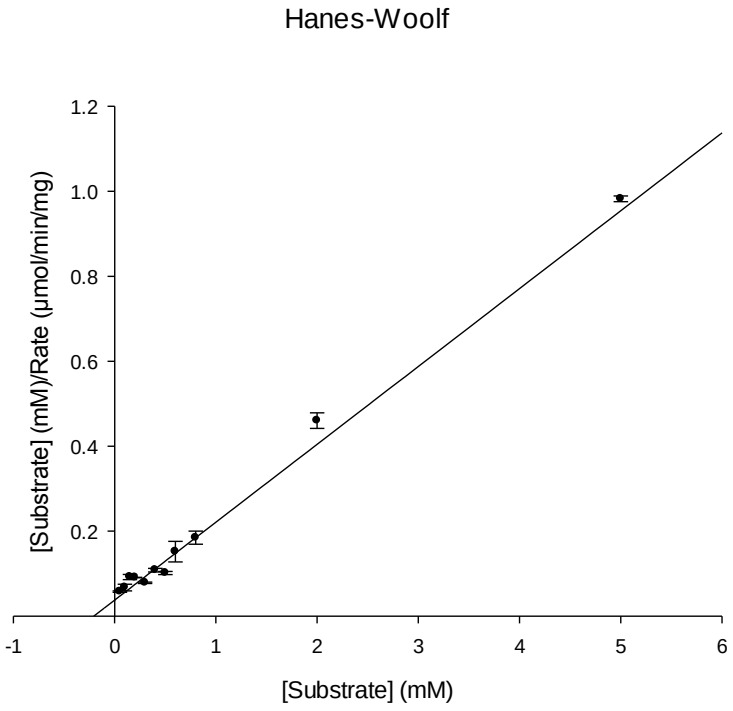


Figure 136 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and 5-norbornene-2-carboxaldehyde in acetonitrile. Conditions: 50mM sodium phosphate pH 7.0, 0.2mM NADPH, 50 °C, 1ml assay volume.

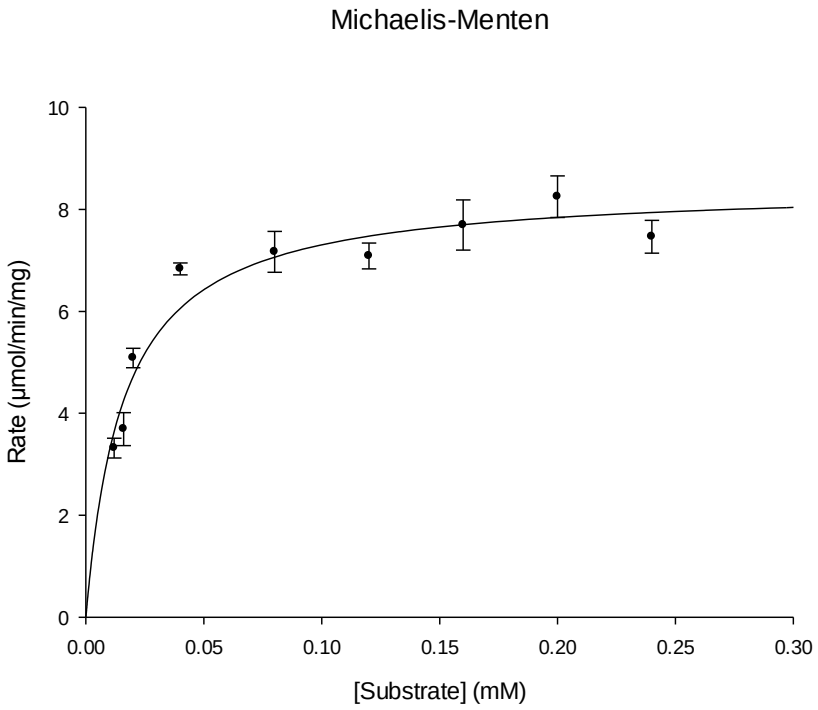


Figure 137 - Graph of rate of reaction vs [S] using ADH G and NADPH. Data fitted to a Michaelis-Menten curve using the Levenberg-Marquardt algorithm by Sigmaplot 12. Conditions: 50mM sodium phosphate pH 7.0, 10mM butanal, 50 °C, 1ml assay volume.

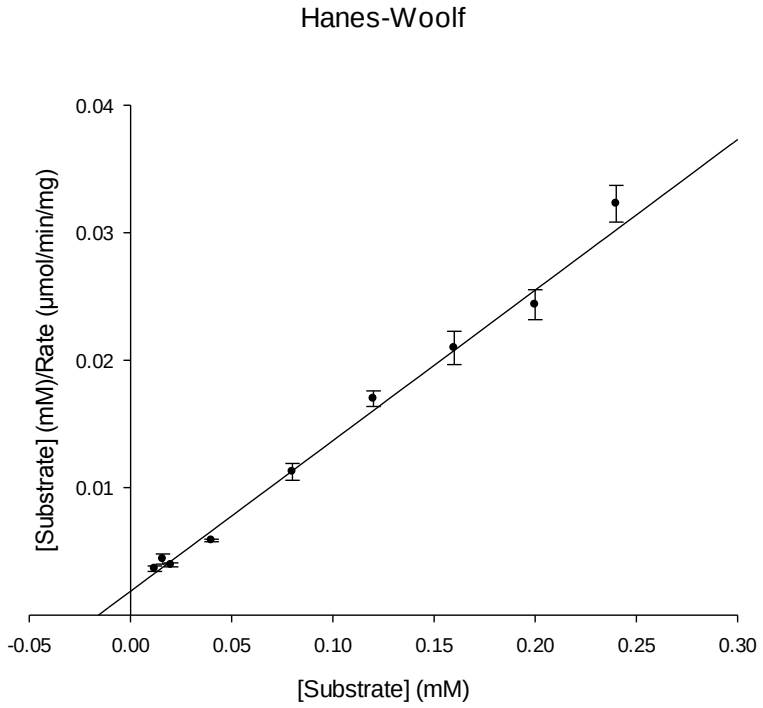


Figure 138 - Hanes-Woolf plot ([S]/v vs [S]) for ADH G and NADPH. Conditions: 50mM sodium phosphate pH 7.0, 10mM butanal, 50 °C, 1ml assay volume.

Substrate	V_{max} (U mg^{-1})	K_M (mM)	K_i (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
Butanal	10.53±0.19	1.88±0.16	-	520	280
Octanal	16.88±1.70	0.37±0.08	5.68±1.59	830	2300
5-Norbornene-2-carboxaldehyde	5.46±0.28	0.21±0.04	-	270	1300
NADPH	8.46±0.25	0.016±0.002	-	420	26000

Table 30 - Summary of kinetic data obtained with ADH G.

As noted with previous enzymes, it appears that ADH G has a preference for longer linear aldehydes despite the introduction of substrate inhibition. When investigating octanal and 5-norbornene-2-carboxaldehyde, both were prepared in stock solutions in acetonitrile. This was necessary to obtain sufficiently concentrated stock solutions; however, even when added to the assay concentrations higher than those reported (5mM) resulted in occlusion of the assay. Thus even when prepared in solvent, upon dissolution into buffer, there was clearly insufficient solvent to allow the substrates to mix fully.

It was expected that being limited to such low concentrations of substrate would substantially impact upon the quality of the data obtained; however, the Michaelis constants of both enzymes were considerably lower than the 5mM threshold, meaning that in both cases the kinetics experiments were in excess of both the $3K_M$ and $10K_M$ points. The only issue resulted in the broad error in calculating the K_i for octanal, which was estimated to be above the range of substrate concentrations tested.

Strictly interpreting the specificity constants, it is not possible to determine a pattern in substrate acceptance based solely upon size of the substrate. Given that octanal is preferred of the three substrates tested then the larger substrate 5-norbornene-2-carboxaldehyde, and then butanal the least, alternative substrates similar to octanal and larger should be tested. In order to effect that, further solvent work is likely to be needed to solubilise the larger substrates in a monophasic system for study at the higher concentrations needed. At a minimum it would be necessary to ensure that 5mM is maintained as a substrate concentration tested, which would prove difficult with larger substrates without higher levels of solvent.

3.4.3. STABILITY ANALYSIS

Given the lack of emphasis with this enzyme, only brief consideration of solvents was done, looking at operational solvents, or those that could be used for preparing stock solutions for substrates in kinetics work.

Volume (vol%)	Activity (DMSO) [%]	Activity (Ethyl acetate) [%]	Activity (Acetonitrile) [%]
1	86	52	99
5	48	116	121
10	45	133	113

Table 31 - Solvent screen, ADH G, 10mM butanal, NADPH.

Acetonitrile was chosen as the solvent of choice, with DMSO being totally unsuitable due to the immediate reduction in activity even with 1vol%. Ethyl acetate was similarly discounted for its inactivating ability, although higher than expected results were obtained at 5 and 10vol% tests. As acetonitrile had little effect at low concentration, and perhaps an enhancing rate at higher concentrations, it was chosen as the best of the three.

When used to prepare stock solutions, both octanal and 5-norbornene-2-carboxaldehyde were prepared at 100mM concentration, meaning that the highest concentration tested, 5mM, corresponded to a 5vol% addition of solvent.

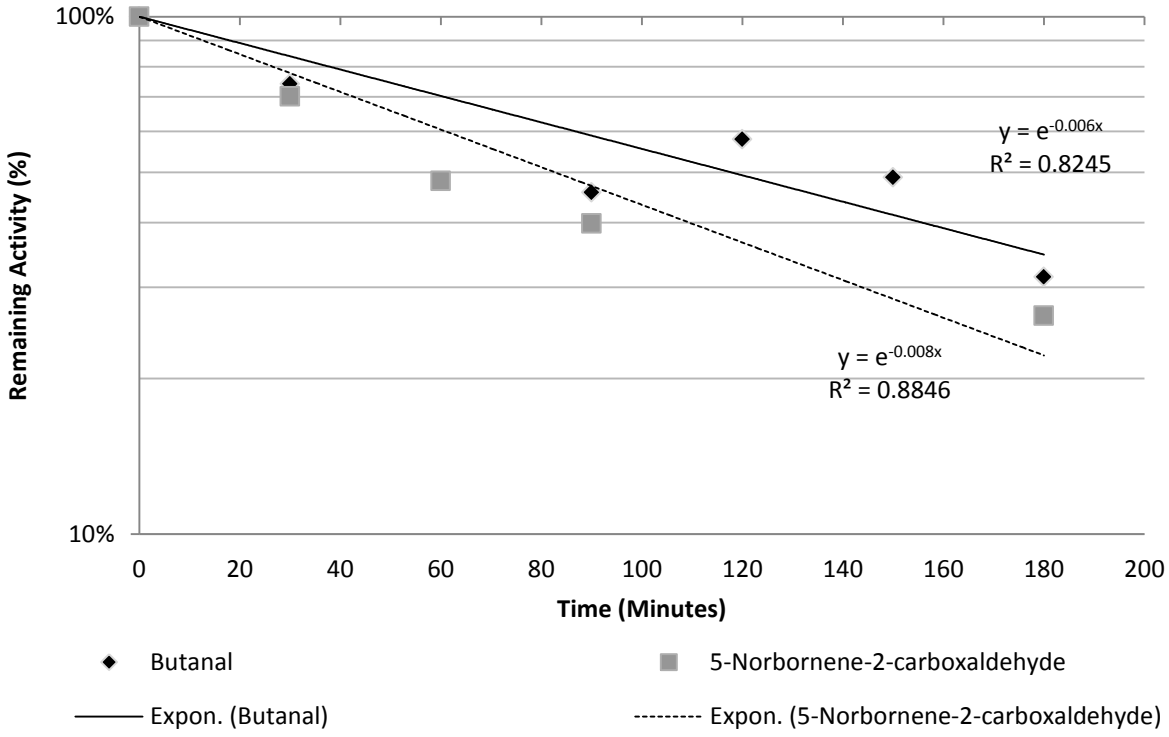


Figure 139 - Thermostability test with ADH G, 60°C, NADPH.

Thermostability of the enzyme was tested twice during the investigation, both over three hours at 60 °C, to determine stability at an operational temperature rather than at elevated temperatures. Given the apparent variance between the two substrates, both rates can be calculated and compared:

$$t_{1/2} = \frac{ln(2)}{0.008} \; or \; \frac{ln(2)}{0.006}$$

$$t_{1/2} = 87 - 116 \; mins$$

It should be noted that the difference in substrate should not affect the inactivation of the enzyme. Therefore the variation noted is likely solely due to the relatively high scatter in the data and experimental variation. The data give a range of 1.5-2.0 hours for the half-life of ADH G. Further analysis of the enzyme, with regards to its optimal temperature, indicated that 60 °C was its optimal temperature, with very rapid decrease in activity above this point, which implies that this enzyme is only stable at the lower end of *G. thermoglucosidasius'* operating temperature.

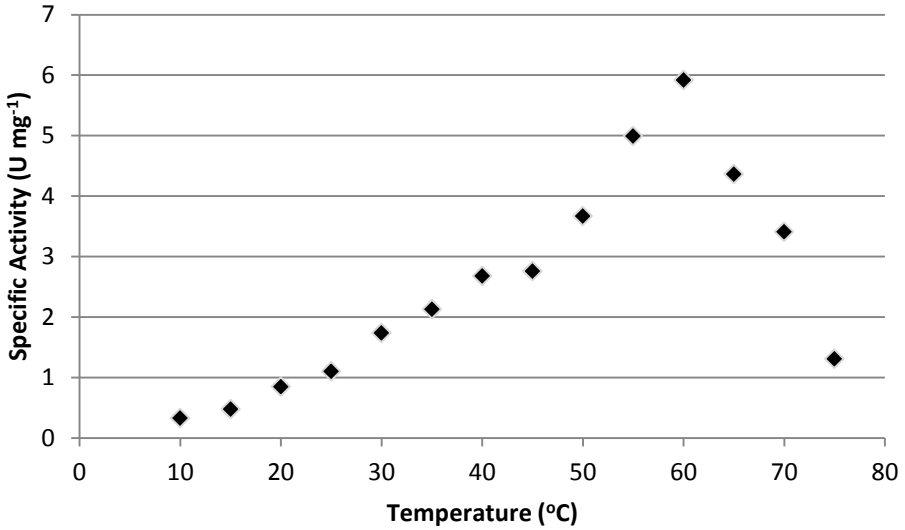


Figure 140 - Optimal temperature study with ADH G, butanal as substrate and NADPH as cofactor.

The thermal shift assay provided two inflection points, at 68.2 °C and 76.9 °C, although the second is very weak compared to the first. This could be indicating a two-step denaturation process or it may be caused by having two different proteins produced, which would require purification to separate the proteins or recloning with the shorter gene. Reanalysis of that protein would provide confirmation of the correct interpretation. It may also be instructive to obtain the larger protein for additional validation, which may be possible through careful purification and isolation. Alternatively removing the internal ribosome binding site from the expanded gene may be preferred.

3.4.4. CHEMICAL ANALYSIS

	Conversion	
Substrate	**1 Hour**	**16 Hours**
5-Norbornene-2-carboxaldehyde	47%	11%
Butanal*	9%	11%
Furfural	28%	Activity
Methylbenzoyl formate	-	-
Decanal	3%	Activity
Benzaldehyde	0%	Activity
4-Chlorobenzaldehyde	10%	12%
2,3-Pentanedione	0%	0%
2,3-Butanedione	0%	0%

Table 32 - Summary of end point assays analysed by GC/MS carried out with ADH G. *Suspected evaporation of substrate

In accordance with the other enzymes, furfural, decanal and benzaldehyde results were merely quantitative in the 16h experiments, which limits the usefulness of the data obtained. It is noticeable though that there is little effect of increasing the time from 1h to 16h in terms of 4-chlorobenaldehyde conversion. This substrate is likely the best example of the substrates tested here, and a fairly rapid denaturation of the enzyme is likely the root cause of this, even at the assay temperature of 50 °C.

3.4.5. SEQUENCE ALIGNMENT

The first striking thing about the MSA is that it appears that ADH G has 34 additional amino acids appended to the N-terminus. This has likely occurred because an early draft of the *G. thermoglucosidasius* genome was misannotated, indicating the gene started from the wrong methionine residue. This mistake was unfortunately incorporated into the gene cloned into *E. coli*, probably along with a ribosome binding site. Thus the two different proteins which were purified are most likely as a result of two different start points for translation.

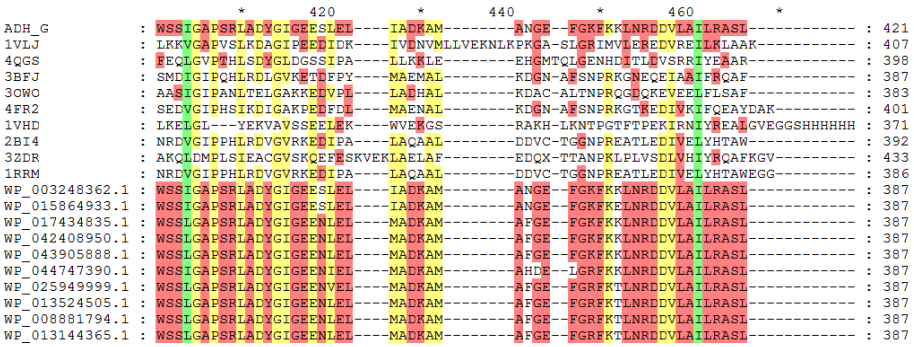


Figure 141 - Multiple Sequence Alignment of ADH G with closest matches from the PDB and in the NCBI reference sequence database (obtained through NCBI Seqr Sequence Search). All proteins annotated as butanol dehydrogenases (BuDH) from the genera *Geobacillus* and *Bacillus*. Crystal structures are annotated as follows: 1VLJ is a BuDH from *Thermotoga maritima*, 4QGS is an aldehyde reductase (YqhD) from *E. coli,* 3BFJ is a 1,3-propanediol oxidoreductase from *Klebsiella pneumoniae,* 3OWO is an iron dependent ADH from *Zymomonas mobilis*, 4FR2 is an ADH from *Oenococcus oeni*, 1VHD is an iron containing ADH From *Thermotoga maritima,* 2BI4 is a lactaldehyde:1,2-propanediol oxidoreductase from *E. coli*, 3ZDR is the ADH domain of a bifunctional ADHE dehydrogenase from *G. thermoglucosidasius*, 1RRM is a lactaldehyde reductase from *E. coli*. MSA was conducted using default MUSCLE settings by MEGA 6.0 (Tamura et al., 2013). Results were visualised using GeneDoc 2.7 (Nicholas et al., 1997) with residues conserved across 100% of entries coloured green, residues conserved across 80% of entries coloured yellow and residues conserved across 50% of entries coloured red. Blue triangles indicate dimer interface residues, purple triangles indicate active site residues and grey triangles indicate residues at the metal binding site. All these sites provided via the NCBI Conserved Domain Database and obtained through the Conserved Domain Search.

1VLJ is a NADH dependent butanol dehydrogenase from *Thermotoga maritima*, but has not been published and thus shall not be discussed further.

4QGS is an apo form of the NADPH dependent ADH, also known as YqhD, from *E. coli*. This enzyme has been extensively studied and is a dimeric ADH of the iron-containing ADH family, although it contains zinc rather than iron in its active site. This zinc ion is co-ordinated by D194, H267 and H281, corresponding to D244, H318 and H332, and matching the notation in Figure 141. The remaining residue H198 (H248) is only co-ordinating in the apo structure; once the cofactor binds the conformational change means the residue forms a hydrogen bond with the cofactor instead. The cofactor binds to two specific GGGS motifs, the first one at positions 135-8, which is absolutely conserved among all sequences. The second motif is at positions 74-78, and this is not conserved amongst all sequences (Sulzenbacher et al., 2004). For example 3OWO does not have this second motif, and has been previously described as a NADH dependent ADH. Thus this motif is key to determining the cofactor specificity of these enzymes. Other key residues for cofactor binding are D140 (conserved), N197 and G199 (not conserved), implying that these residues are also associated with cofactor specificity (Sulzenbacher et al., 2004). Experimental evidence has also been obtained, with double mutant D140Q N197H both decreasing K_M values and increasing k_{cat} values with some substrates (Jarboe, 2011).

Despite extensive testing with a variety of alcohols, aldehydes, amino acids and sugars, activity was only originally noted with a small range of alcohols. Whilst these range from 1-propanol to benzyl alcohol, in no case was a V_{max}>5 min^{-1} reported, and K_M values ranged from 5-35 mM (Sulzenbacher et al., 2004). Subsequent investigation has provided additional information, with small aldehydes proving to be better substrates. With K_M values as low as 0.4 mM, and k_{cat} values as high as 60 s^{-1} butanal appears to the best substrate tested. Activity was also noted with glyceraldehyde, isobutyraldehyde and furfural among other substrates (Jarboe,

2011). This range of substrates has led to YqhD being utilised in a variety of metabolic engineering projects, but most notably isobutanol and other biofuel production (Ran et al., 2008; Atsumi et al., 2010; Jarboe, 2011; Pick et al., 2013). It is also notable that suppression or deletion of this enzyme confers furfural resistance to *E. coli* as the conversion of furfural to furfuryl alcohol consumes enough NADPH that the cell is overwhelmed; in particular cysteine biosynthesis is stalled. One last key point is that YqhD has been suggested to have interfered with previous investigations into heterologously produced ADHs in *E. coli* due to its relatively low K_M with butanal (0.5 mM). High background rates have been noted and several ADHs studied over the last decades may have been affected (Jarboe, 2011).

All other crystal structures have been mentioned previously. 3BFJ is a NADH dependent 1,3-propanediol oxidoreductase from *Klebsiella pneumoniae*, a member of the iron-containing ADH family. The residues annotated as co-ordinating the iron ion are the same ones that are annotated and are those conserved throughout the MSA. The most obvious difference is that 3BFJ is NADH dependent rather than NADPH dependent. As mentioned, this appears to be associated with the loss of the GGGS motif at positions 74-78, and it is not present in 3BFJ.

3OWO is an iron-containing ADH from *Zymomonas mobilis*. It is NAD^+ dependent, oxygen sensitive, dimeric and its native role is to convert ethanal to ethanol to provide NADH as part of the cytosolic respiratory system. In place of the NADPH associated residues mentioned above, there is D39 (D71), a key residue to ensure NADH dependence. Unlike ADH E, ADH G does not have the same residues noted with 3OWO to regulate substrate specificity, F149 and F254 (F199 and F307).

Lastly, 4FR2 is an ADH from *Oenococcus oeni.* A dimeric enzyme that is NADH dependent, it shares many of the same features as 3OWO. Although a member of the iron-containing ADH family, the active site contains nickel and is entirely inhibited by zinc. The enzyme can catalyse a wide range of small aldehydes and alcohols including 1,2-propanediol, 1,3-propanediol, glycerol, butanal and 2-propanol (Elleuche et al., 2013).

To conclude, this MSA appears to have both NADH and NADPH dependent enzymes as the most similar to ADH G. Fortunately it appears to be easy to determine which cofactor a given enzyme would be able to accept, due to easily identifiable and conserved residues. That said, there are a very large number of residues annotated as being relevant to the active site, dimer interface and co-ordinating the metal, and it is only the latter residues that are all conserved. The majority of residues annotated as being important in the active site are only conserved amongst the most similar proteins and not the crystal structures. The presumption is that these residues are related to the substrate specificity of the enzymes. The residues involved with cofactor binding have been identified and are conserved entirely, yet both dimer residues and activity residues, despite annotation, there is no definitive evidence to prove their cases.

3.4.6. SUMMARY

ADH G is a relatively insignificant enzyme in terms of its industrial relevance. A broad spectrum aldehyde reductase with relatively low activity, its general weakness at higher temperatures and variable reaction to solvents would seem to remove any particular advantage of this enzyme over any other. That said, it does have potential in converting particularly complex aldehydes, and therefore should be considered for such purposes if a recalcitrant substrate should appear. It is a certainty that the full substrate scope for this enzyme has yet to be elucidated.

4. DISCUSSION AND FURTHER WORK

The results obtained in this work will be discussed in further detail in this section, with reference to their biological relevance and potential industrial significance. Where additional work is likely to be required to facilitate a better understanding this will be clearly indicated.

4.1. RESULTS SUMMARY

All specific activities obtained with alcohol dehydrogenases in this study are compiled in Table 33 to better enable comparisons to be made. In broad terms a typical specific activity obtained is in the region of 1-100 U mg^{-1}, with only ADH E exceeding this with specific activities recorded of nearly 400 U mg^{-1}. Notably, however, ADH E had a much more restricted substrate scope compared to all other enzymes in the table.

Substrate	ADH B	ADH D	ADH E	ADH F (NADH)	ADH F (NADPH)	ADH G
Methanal		0.9		0.1	0	0.4
Ethanal	56		362			
Propanal	79		376	0.7	0	2
Butanal	95	10	227	1	5	8
Pentanal	30					
Hexanal	0.8		99	5	0.6	9
Octanal			280	7	2	12
Decanal		4	9	0.3	2	7
Isobutanal		9	7	1	5	0.5
Glyoxal		12	0	12	14	3
3-Methyl butanal		0	0			
2-Butanone	76	0	0	0	0	0
2-Pentanone	25					
2-Hexanone	2					
Acetoin	33	0				
2,3-Butanone		13	0	16	24	
3-Methyl-2-butanone			0	0	0.1	0
2,3-Pentanone		6	6	14	11	0
3-Pentanone		0	0	0	0	0
2,4-Dimethyl-3-pentanone		0				
2,3-Hexanedione		11	4	20	22	
2,5-Hexanedione		0	0	0	0	
3,4-Hexanedione		3	3	8	2	
2,3-Heptanedione		11	3	21	33	
3-Hexanone	0.4					
2-Decanone		0	0	0	0	0
Cyclohexanone		0	0	0	0.9	0
Furfural		7	122	3	9	5
Acetophenone		0	0	0	0.3	0
Methyl acetoacetate		0.4		0.6	0	
Ethyl acetoacetate		0.6	0	0.1	1	0
Benzaldehyde		7	3	5	8	2
4-Chlorobenzaldehyde		19	0	5	4	5
Pyruvate		0	0	0	0.1	0
Ethyl pyruvate		3	0	6	3	0

Ethyl levulinate		0	0	0	0.1	0
Levulinate		0	0	0	0	0
α-Tetralone		0	0	0	0	
Benzophenone		0		0	0	
Benzoin		+		0	0	
Benzil		0		0.3	0	
5-Norbornene-2-carboxaldehyde		13	0	10	7	7
1-phenyl-1,2-propanedione		11	0	21	11	0
Ethyl 2-chloroacetoacetate		9		12	11	
Methyl 4-chloroacetoacetate		3				
Ethyl 4-chloroacetoacetate	91	5	0	5	2	0
Ethyl 4,4,4-trifluoroacetoacetate	0.1	0	0	0.7	3	0
6-Methyl-5-hepten-2-one	0					
4-Phenyl-2-butanone	0	0	0	0	0.3	0
2,2,2-Trifluoroacetophenone	0	0	0	3	9	
Ethyl-2-oxo-4-phenylbutyrate	0.9	7	0	2	4	0
2',3',4',5',6'-Pentafluoroacetophenone		0	0	0.3	3	0
2,2',4'-Trichloroacetophenone				2	0	
4'-Chloroacetophenone				0.2	0	
4'-Bromoacetophenone				0.4	0	
2-Chloroacetophenone				0.2	0	
Propiophenone				0	0	

Table 33 - Consolidated results of specific activities noted with all ADHs studied. No specific activities available for ADH A or ADH C; gaps indicate substrate not tested. + Unquantified Activity. All figures given in U mg^{-1}

Due to the method of screening, involving a common primary screen and an *ad hoc* secondary screen based upon each enzyme's activities, it has been possible to make comparisons between each enzyme. For example, where an ADH has shown activity with methanal, it is always relatively low, as this substrate is known to be particularly harsh on enzymes. In the case of the industrially related substrates, with the exception of ethyl 4-chloroacetoacetate, all the highest specific activities are noted with ADH D and ADH F, which is why these two enzymes were particularly heavily studied.

4.2. NATIVE ROLE FOR ALCOHOL DEHYDROGENASES

The first two ADHs that will be considered are ADH D and ADH F as they were the most heavily studied in this investigation. Both are annotated on the National Center for Biotechnology Information (NCBI) database as Aldo-Keto Reductases, with the former being described as an aryl-alcohol dehydrogenase, and the latter as a methylglyoxal reductase. With estimated sizes of 35.5kDa and 31.5kDa (315 and 275 residues) respectively and with both enzymes being heavily conserved throughout the genera *Geobacillus, Bacillus, Anoxybacillus,* and many other related genera, it is likely that the annotation is at least approximately accurate.

That said, there is certainly similarity to other AKR families, with ADH D and similar proteins being annotated as a "Voltage Gated Potassium Channel Beta Subunit" as well, both on the NCBI database and elsewhere, which is another AKR family. ADH F is not similarly affected, with annotation referring either to glyoxal reductase or aryl-alcohol dehydrogenase.

Whilst the assumption could be made that ADH D and ADH F are similar enzymes, which their substrate profiles would suggest as much as their annotations, this would not be correct. Their sequence similarity is very low (26% identity) and thus despite their apparently similar

substrate profiles are unlikely to perform similar functions *in vivo*. Noting that ADH F is similar to the *Bacillus subtilis* AKRs YvgN and YtbE (67% identity), which have been previously studied (Lei et al., 2009), then the *B. subtilis* putative functions should be considered. YvgN is believed to be involved in sporulation, whereas YtbE is known as essential to *B. subtilis* for the removal of toxic aldehydes (Commichau et al., 2013). Both, however, catalyse the reduction of aldehydes to their respective alcohols.

ADH D is expressed natively at reasonably high levels, 0.099% of total RNA under anaerobic conditions and 0.066% under aerobic conditions; ADH F is barely expressed at all, 0.016% of total RNA under aerobic conditions and 0.012% under anaerobic conditions (Leak et al, unpublished). A low but constant expression of ADH F would be consistent with detoxification, and high expression of a putative sporulation enzyme is somewhat inconsistent. Whilst *G. thermoglucosidasius* does form spores, it is only known to do so under stress, thus it seems more likely that ADH F has a function similar to *B. subtilis* AKR YtbE, although more work is likely to be needed to conclusively prove this.

Consulting the placement of the gene encoding ADH D, it appears to be the first gene in a 9 gene cluster or operon annotated with reference to sugar transport and particularly mannitol. Notably, another alcohol dehydrogenase is incorporated into this cluster as well. As AKRs are also known for aldose reduction as well as alcohol reduction, this could imply that ADH D is natively related to sugar transport and utilisation. This cannot be refuted or confirmed by the substrates investigated as alcohols and sugars were not tested. Therefore the first steps towards confirming this hypothesis should be to test ADH D with a range of new substrates, specifically mannitol, mannose and related hexoses, to determine activity. If confirmed, gene deletion may provide an insight into the native role of ADH D. As *G. thermoglucosidasius* is able to utilise both mannitol and mannose (Fong et al., 2006), a strain deficient in ADH D should have a slower growth rate than normal or be unable to grow entirely when grown exclusively on these carbon sources.

It is less likely to be possible to determine the detoxification effects of ADH F by simply removing the gene as with ADH D. As demonstrated with the similar substrate profiles of ADH D and ADH F, even if ADH F were removed, there will be an alternative enzyme able to replace it.

The other ADHs studied have a variety of annotations. ADH A is well known as the *G. thermoglucosidasius* variant of the AdhE bifunctional enzyme, ADH B is annotated as a threonine dehydrogenase, zinc-dependent alcohol dehydrogenase and a member of the MDR superfamily, ADH E is indicated as a bifunctional ADH/aldehyde dehydrogenase and a member of the iron containing ADH superfamily, and ADH G is annotated as an NADH dependent butanol dehydrogenase and member of the iron containing ADH superfamily.

Turning first to gene placement, the gene encoding ADH B is not associated with any further genes that would provide an indication as to the *in vivo* role of the enzyme. It is of note that it is very highly expressed under aerobic conditions, but almost entirely suppressed under anaerobic conditions, with 0.338% of total RNA and 0.009% respectively. This could be an indication that ADH B is directly involved with primary metabolism, specifically aerobic metabolism. As ADH B utilises NADH as a cofactor and is annotated as a zinc containing member of the MDR family, then this implies that it has dehydrogenase activity *in vivo*. From the substrates tested this would mean that either 2-butanol or 2-propanol are the preferred substrates of this enzyme.

The position of the gene encoding ADH E would support the hypothesis that this enzyme is involved in propanal removal, being in a 15 gene cluster annotated as involved in converting

1,2-propanediol to 1-propanol via propanal, as well as the utilisation of ethanolamine. During the course of this investigation it has been shown that ADH E has a very high activity with both propanal and ethanal. Therefore whilst propanal degradation is the most likely *in vivo* purpose, this does not preclude the possibility there ADH E is also involved in ethanolamine degradation as well. It appears to be very poorly expressed under aerobic or anaerobic conditions, with approximately 0.005% of total RNA in both conditions.

In the case of ADH G, its gene is adjacent to a number of sugar transporters, a gene annotated as coding for glucose-6-phosphate isomerase and a lac repressor gene. This implies that ADH G is involved with the glycolysis pathway in some manner, although the exact manner of this relationship is difficult to say without additional work. One of the most similar available crystal structures refers to an *E. coli* ADH designated YqhD, which has activity with a variety of small aldehydes and alcohols, which is similar to the noted activities of ADH G. It has also been determined that YqhD is required for 1,3-propanediol production, and has been used to produce 1,2-propanediol (Jarboe et al., 2010), but these are not necessarily natural reactions for this enzyme. It is proposed that the native substrate is an alcohol longer than C3 (Sulzenbacher et al., 2004), but this is only based upon kinetics. Therefore at this time it is not possible to definitively identify the *in vivo* role for either YqhD or ADH G. Whatever the exact role performed by the enzyme, it is expressed moderately well in aerobic conditions (0.023%) and slightly better under anaerobic conditions (0.036% of total RNA).

4.3. INDUSTRIAL RELEVANCE OF ENZYMES

Again considering ADH D and ADH F first, both have substantial industrial relevance in terms of their substrate profiles and other features. Taken together, their substrate profiles include aldehydes and ketones of significance including:

Substrate	Industrial Use
5-Norbornene-2-carboxaldehyde	Polymers
1-Phenyl-1,2-propanedione	Pharmaceuticals
Ethyl 4-chloroacetoacetate	Statins
Ethyl-2-oxo-4-phenylbutyrate	Pharmaceuticals
2,2,2-Trifluoroacetophenone*	Liquid Crystals
2',3',4',5',6'-Pentafluoroacetophenone*	Pharmaceuticals

Table 34 - Industrially relevant substrates that ADH D and ADH F have demonstrated activity with. *ADH F only.

Both ADHs D and F demonstrated potential with the above substrates; however, issues were noted with the chemical analysis, attributed to cofactor degradation. For the purposes of analysis, this could be avoided by increasing the enzyme volume used, to ensure that the reaction completed ahead of the cofactor degrading. Whilst it is highly likely that the cofactor is the issue, it is worth confirming this before committing to any further large scale work by conducting an assay over several hours, adding further cofactor on an hourly basis. This should prove that cofactor degradation is the root cause of the issue. Irrespective of which component degrades, additional enzyme added at the start of the reaction should allow for a final conversion to be determined.

It is notable that ADH D was able to convert 95% of the substrate 5-norbornene-2-carboxaldehyde, and ADH F also performed well with this substrate, converting 78% after one hour. ADH F is especially likely to benefit from improving cofactor stability, particularly with NADH, due to the high K_M value noted. The relatively high conversion rates for 5-norbornene-

2-carboxaldehyde give confidence that this is a fruitful area for further work, and should be prioritised to see if other substrates can be catalysed to such a high level. Continuing to work with a two-enzyme, two-substrate based system with a GDH will ensure that the reaction can be forced to completion, which should rule out issues with the position of equilibrium.

A further industrial consideration is cofactor regeneration itself. Presuming that cofactor degradation is minimised, perhaps by a change of buffer, reduction in temperature of reaction, or via stabilising additives, then the regeneration of the cofactor is also an issue. Whilst using a GDH is acceptable at bench scale, continued use at pilot scale or above is likely to be problematic. Although the continual decrease in pH can be mitigated by effective monitoring of the bioreactor and the gradual addition of sodium hydroxide or similar alkaline, this represents a cost. It will also increase the ionic strength of the buffer, which may cause issues with the reaction – this can be tested if it is the preferred route. An alternative approach would be to utilise a second ADH to regenerate the cofactor with either isopropanol or ethanol as a sacrificial substrate. Recommendations may include ADH B and ADH E, should further testing indicate that either enzyme is suitable for this task. ADH E would need to be tested with alcohols; however, in principle both could be used to regenerate NADH, and ADH E may be useful in regenerating NADPH.

As both ADHs D and F converted 1-phenyl-1,2-propanedione into two products, 1-hydroxy-1-phenylpropan-2-one and 2-hydroxy-1-phenylpropan-1-one, this may be problematic for industrial use. The exact proportions of these products can only be confirmed once work has been done to ensure that the reactions run to completion or as far as can the reaction can be pulled via the GDH. It may prove that for such substrates these enzymes are simply unsuitable without engineering the enzyme to manipulate the conversion ratios but this will entail a substantial period of work including obtaining a crystal structure of the enzyme and determining how the substrates are catalysed. Site saturation mutagenesis of target residues in the active site and further assaying with the substrates may then result in a mutant with the ability to convert 1-phenyl-1,2-propanedione into the required product.

Following on from the issue of multiple products from catalysis of a diketone, the next stage of chemical analysis with these ADHs will be to determine if they produce enantiopure products. As chiral synthesis is the major advantage of utilising enzymes over chemocatalysis, aside from environmental consideration, then it is of critical importance to determine the enantiomeric excess of the products formed.

ADH D has demonstrated substantial solvent resistance, with no negative effects noted with up to 20vol% of acetonitrile, ethyl acetate and CPME. ADH F is less tolerant of solvents, but still retains substantial activity in the presence of the same solvents. Interestingly, it appears to be resistant, or perhaps activated, by low levels of hexane. Therefore there is a *prima facie* case for industrial processing in the presence of solvents for both enzymes, which are likely to be required in order to facilitate high concentrations of substrate and product.

Due to the aforementioned high concentration of substrate necessary for a viable industrial process, it is necessary to ensure that enzymes are able to tolerate higher levels of substrate and product. There are alternative methods, such as immobilising enzymes or operating in fed batch mode, to avoid excessive substrate concentrations; however, it is generally easier to ensure that the enzymes to be utilised in an industrial process are able to cope with expected concentrations of substrate and product. No substrate inhibition was detected when working with ADH D, and ADH F only demonstrated substrate inhibition in rare cases, with estimates of K_i typically falling either on the border of substrate concentrations tested, or outside the range used. To elucidate whether the enzymes are suitable for a full scale industrial process then substrate concentrations of at least 100mM, and preferably >500mM should be tested. It is

unlikely that testing such concentrations will be simple, due to hydrophobicity of the larger substrates. Through judicious use of relevant solvents it may be possible to reduce or eliminate substrate inhibition entirely by keeping to a biphasic system; for example, whilst the substrate and product may have solubilities in excess of 1M in hexane, even with 20vol% solvent, there may be a very limited substrate concentration exposed to the enzyme in the aqueous phase. Typically this is how such substrates are catalysed by isolated enzymes in industrial processes.

If the above recommendation of using ADH E to regenerate the cofactor is explored, then it may be possible to utilise the sacrificial substrate as a cosolvent to create a monophasic environment. In the case of ADH E, it would have to be ethanol as solvent, and would necessitate testing both ADH E and the primary ADH (either D or F) to determine whether such a system was stable. If so, this could be a particularly beneficial use of operating at thermophilic temperatures, as the ethanal by-product can be driven off very easily and even collected for further use, allowing the reaction to be pulled to completion whilst also easing purification downstream.

With reference to the other ADHs studied, ADHs B and E are likely to have usefulness as cofactor regeneration systems far beyond solely with other *G. thermoglucosidasius* enzymes. In particular ADH E, with very high activity when used with NADH, should be tested for activity in the reverse direction. Ethanol or propanol as cosolvents and sacrificial substrates would be particularly beneficial due to the boiling points of ethanal and propanal of 20 ^{o}C and 50 ^{o}C respectively. As the putative *in vivo* role for this enzyme is as a dehydrogenase, operating conditions may need to be adjusted to ensure a high rate of reaction in the appropriate direction for cofactor regeneration.

Other potential industrial uses for ADH B would be in the production of statins, through the utilisation of ethyl 4-chloroacetoacetate as substrate. Due to ADH B also accepting small alcohols, it may be possible to establish a two-substrate one-enzyme cofactor regeneration system by using ethanol, isopropanol or similar alcohol as cosolvent. The most efficient method for preparing such a system would be to determine the resistance of ADH B to using 10vol% or 20vol% alcohol as a cosolvent, and establishing whether production of ethyl 3-hydroxybutyrate can be maintained in the presence of high quantities of sacrificial alcohol.

Further afield, both ADH B and E have demonstrated potential in the area of bioremediation due to their high activity with a range of small substrates. There are a range of unusual industries with issues in dealing with volatile aldehydes and ketones including the aluminium industry (Laird et al., 2013), human waste management (Lin et al., 2013) and explosives remediation (Singh et al., 2008); wherever volatile aldehydes and ketones pose a risk to health or the environment, alcohol dehydrogenases may be considered as a potential mitigation step. In many instances recovery of these volatile organic compounds would be necessary, or some kind of *in situ* system designed to allow biocatalysis to be utilised. This is not a trivial step to take, and would necessitate enzyme or organism lifetime to be determined. Utilisation of enzymes in such environments is likely to be a relatively long term objective.

Whichever field the ADHs may be suitable for, once a particular industrial process or task has been identified as appropriate for an ADH, the next stage of the feasibility study will be to utilise model reactions as appropriate for that process. This is likely to involve scaling up reactions substantially, to produce grams of product at a time, necessitating larger amounts of enzyme.

In order to produce such large quantities of enzyme in a timely fashion, optimisation of the production and purification method is required. This was not carried out as sufficient enzyme

was always produced for study. Typical parameters to consider in expression might be: IPTG concentration, length of time prior to and post induction, media composition and temperature of production. Alternative approaches may include considering promotor modifications or attempts at extracellular production and attempts to reduce inclusion body formation.

From the purification side, it may be expedient to continue using IMAC, but it is likely that future work will have to consider alternative methods of purification based upon the enzyme's properties, and how this may affect the stability of the enzyme.

4.4. ADDITIONAL AREAS FOR INVESTIGATION

Given ADH F is annotated as an AKR and has substantial activity with NADH, then this enzyme has substantial biological significance and should be subject to further investigation as activity with NADH is relatively rare (Di Luccio et al., 2006). In particular, the full potential of ADH F using NADH as a cofactor should be examined, as it seems likely from work conducted thus far that higher activities may be possible with NADH rather than NADPH.

In addition, from previous work undertaken with the *B. subtilis* AKRs, it seems likely that the preferred substrates of ADH F have not yet been identified (Lei et al., 2009). Comparing with the range of substrates tested in this work, all that can be said is that the preferred substrates analysed in both cases have kinetic constants of similar magnitude. As completely different substrates were chosen for ADH D and F compared to YtbE and YvgN, it may be productive to consider screening both ADHs for activity with the same range of aldehydes, e.g. 2-chlorobenzaldehyde and 2,4-dichlorobenzaldehyde, and consider analysing the kinetics with substrates that have already been demonstrated to have activity with ADH D and F, such as benzaldehyde, 4-chlorobenzaldehyde and glyoxal. Not only is determining the substrate scope useful in and of itself, determining whether either ADH D or ADH F have identical or differing substrate promiscuity to *B. subtilis* enzymes with known activities would be enlightening. As thermophilic AKRs are relatively understudied at present, with few present in the literature, comparisons between mesophilic and thermophilic AKRs in similar genera should be a target (Drury and Penning, n.d.; Willies et al., 2010).

With >60% amino acid identity to *B. subtilis* AKRs previously mentioned, this means ADH F can be tentatively assigned the designation AKR5G4, family 5, subfamily G (Drury and Penning, n.d.). Due to low sequence identity, ADH D is harder to place within the AKR superfamily; however, by annotation it is likely to fall within family 6 with the other potassium channel proteins, tentatively assigned as AKR6D1. Accurately placing these thermophilic AKRs within the wider AKR superfamily would allow for further comparisons with similar mesophilic enzymes and therefore should be considered.

Stepping back from concentrating on ADHs A through G, other enzymes have been isolated during this investigation that due to time constraints, have not been considered. In total, eighteen putative ADHs were investigated to some degree. All those not previously mentioned have been isolated successfully, but have yet to be assayed.

4.5. SUMMARY OF IMPLICATIONS FROM SEQUENCE ALIGMENTS

There are some additional points raised from the sequence alignments which may inform the direction of further work:

- ADH B may have no enantiospecificity, and it may also be possible to raise the T_m of ADH B by a changing a single amino acid.
- ADH D may be the founding member of a new subfamily of AKRs, and the enzyme may have an *in vivo* role of cofactor recycling rather than the production of aldehydes, ketones or alcohols. A crystal structure is likely to be particularly helpful in determining the likely *in vivo* role, as is investigation of the *G. thermoglucosidasius* sugar uptake system.
- ADH E has very few highly similar sequences available, and therefore it may be useful to obtain a crystal structure of this enzyme as well. This is reinforced by the substantially different activity recorded with ADH E and its most similar enzymes. It may be possible to alter the co-factor specificity of ADH E relatively easily by altering one or few amino acids. If proven then this could allow the enzyme to be used as a thermophilic cofactor regeneration system. The presence of a metal ion in ADH E may also be of substantial interest, given the range of ions in similar enzymes.
- ADH F may have significant industrial relevance as a similar enzyme has already been considered in the production of pharmaceutical intermediates, yet lacks stability at the necessary temperatures.

These points should be considered in conjunction with the areas previously discussed.

5. METHODS

For clarity, the methods used have been split into chronologically arranged sections. In addition, where relevant, distinction is made between methods used in Chapter 2 and those used in Chapter 3.

5.1. MOLECULAR BIOLOGY

All genes encoding the enzymes under study were identified using publicly and privately held genome sequences of *G. thermoglucosidasius* with the privately held sequence taking precedence. The publicly available sequence is held by the National Center for Biotechnology Information and is *G. thermoglucosidasius* C56-YS93. The private genome, TM242, is owned by TMO Renewables, and held by RAST, Rapid Annotation using Subsystem Technology (Aziz et al., 2008). Throughout this thesis, internal designations have been used to avoid confusion, and a full list of all designations is available in Section 7.1.

Suitable primers were designed by standard methods, aiming to create as lengthy primers as possible to ensure their T_m was approximately 72 ^{o}C. Dimers and cross dimers were restricted and hairpins severely restricted by the use of introduced silent mutations and by altering the six nucleotides appended to the beginning of each primer. Silent mutations were added wherever possible by referral to an estimated codon bias table for *E. coli* such that any altered codons were at least as common in the species as not to affect expression adversely.

In brief, the primer design process was as follows:

1. It was determined whether restriction enzymes cut through a gene of interest, using NEBcutter v2.0 (Vincze et al., 2003)
2. Using Artemis (Rutherford et al., 2000) to visualise the sequence, an initial sequence for primers was created, using the generic schematics:
    a. Forward Primer: [NNNNNN][*Nhe*I Site][First 30 bases of gene, skipping initial start codon]
    b. Reverse Primer: [NNNNNN][*Xho*I Site][Reverse complement STOP (TAA)][Reverse Complement last 30 bases of gene]
3. Dimers, cross dimers and hairpins were analysed using NetPrimer (Premier Biosoft). If dimers rated above 8 kJ mol^{-1}, cross dimers rated above 8kJ mol^{-1} or hairpin rated above 3kJ mol^{-1}, then silent mutations were added to ameliorate the situation.
4. Manual iterations were conducted until the above conditions were satisfied. In addition, the size of primers was increased or decreased as necessary to ensure that the estimated T_m was approximately 70 °C.

A list of primers used is in Section 7.2, and in all cases presented the restriction enzymes utilised were *Nhe*I and *Xho*I, with the plasmid pET28a used. *E.coli* cloning strains JM109 and BIOblue were used, and the expression strain BL21 (DE3) pLysS was used, with the exception of ADHs 2 and 4, where BL21 (DE3) was used instead.

Polymerase chain reactions were carried out using genomic DNA from the TM242 strain of *G. thermoglucosidasius*, extracted using a soil microbe DNA miniprep kit (Zymo Research), and per manufacturer's protocols. The PCR protocol was adapted based upon the T_m of the individual primers designed; however, the overall method is as follows:

1. Initial Denaturation Step: 98 °C for 60 seconds
2. Denaturation Cycle: 98 °C for 20 seconds
3. Anneal Cycle: 70 °C for 20 seconds
4. Elongation Cycle: 72 °C for 60 seconds
5. Termination Step: 72 °C for 5 minutes

The components for a typical PCR were:

- 10µl High Fidelity Buffer
- 1µl dNTPs
- 2.5µl of each primer (from 10x diluted stocks)*
- 1µl genomic template**
- 32.5µl ultrapure water (Milli-Q)
- 0.5µl Phusion enzyme

*Primers were obtained from Eurofins desiccated, and prepared as per their instruction. A 10x dilution was taken as a working primer stock. ** In repeat PCR 0.5µl of previous PCR product or digested DNA has also been used.

This was run using Phusion High Fidelity DNA Polymerase (NEB), and typically 20-30 cycles. The annealing temperature was initially determined by the lowest primer T_m +3 °C, or 72 °C, whichever was lower. This temperature was the only one changed, and was occasionally varied depending on PCR success. With particularly intractable examples, a gradient PCR was applied, whereby the anneal temperature was varied whilst maintaining the same temperatures in other stages. This made it possible to test multiple temperatures, perhaps varying as much as 20 °C, at one time. After this, typical anneal temperatures were markedly dropped to approximately 60 °C, which proved to be more effective than previously recommended temperatures.

PCR products were then analysed by standard DNA gel electrophoresis, using gels made from 70ml TAE buffer (Tris-acetate-EDTA buffer with final concentrations of 40mM Tris, 20mM acetic acid and 1mM EDTA), 1wt% agarose. DNA was visualised using by adding 3.5µl Sybrsafe (Thermofisher Scientific) to the molten gel and UV light. Relevant portions of the gel were excised and the DNA was purified using the Wizard Quick Protocol for PCR clean up by centrifugation, utilising the relevant kit.

The PCR products were then analysed by Nanovue to determine the exact concentration of DNA present, which was then digested for one hour at 37 °C. A standard 30µl digest would comprise 24.5µl of the PCR product or host plasmid, 3µl of the relevant buffer, 0.5µl Bovine Serum Albumin if required (not needed after the manufacturer released new buffers), and 0.5µl of each restriction enzyme, typically *Nhe*I and *Xho*I (both NEB). Typically approximately 1µg of DNA was added to each digest or 24.5µl if the concentration was insufficient.

Post digest, the DNA was purified, following the Wizard Quick Protocol again. Ligation was carried out overnight at 16 °C or for one hour at room temperature. Typical ligations were carried out at a volume of 30µl, using 3µl T4 DNA ligase buffer and 1µl T4 DNA ligase. The remainder is comprised of PCR product digest and plasmid digest; however, the exact volumes of either depend on the concentration of DNA. The concentrations of each were determined by Nanovue and ratios of insert to plasmid varied considerably, with ratios of 6:1 to 1:3 typically used.

Ethanol precipitation was the next step, with all components kept on ice prior to and during this step. To each ligation 3µl 3M sodium acetate (pH 5.2), 82.5µl 100% ethanol and 1µl

dextran blue dye was added. After gentle vortexing to mix all components and left on ice for 30 minutes, each precipitation was centrifuged at 14000 rpm in a VWR microstar 17 centrifuge at 4 ^{o}C for one hour. The supernatant was then gently removed and 100µl 70% ethanol was then gently added and removed immediately, without disturbing the DNA pellet. The pellet was then dried at room temperature for 15-30 minutes, as necessary to ensure complete evaporation of the ethanol.

Transformations were prepared by suspending the plasmid in 20µl Milli-Q water, vortexing very gently as necessary. Electrocuvettes and *E.* coli JM109/BIOBlue samples were buried entirely in ice for at least 30 minutes and LB media was prewarmed at 37 ^{o}C for the same duration. When all components were ready 5µl of DNA was added to an aliquot of *E. coli* and added directly to the chilled electrocuvette and electroporated at 1.8kV using a Genepulser XCell (Biorad). Expediently as possible, 1ml of the prewarmed LB media was then used to wash out the transformed *E. coli* into a 50ml falcon tube. The tubes were then incubated at 37 ^{o}C, shaken vigorously throughout.

After the recovery period 100µl of the transformants would be directly spread to an LB plate containing relevant antibiotics (typically kanamycin) and the remainder briefly centrifuged (8000 rpm for 5 minutes), resuspended in 100µl of the drawn off supernatant and plated onto a second LB plate. Both plates were incubated statically overnight at 37 ^{o}C, perhaps left as long as 24-30 hours to encourage larger colonies.

Should transformant colonies grow, then they would be individually picked using a sterile pipette tip, carefully transferred onto a fresh plate, with the tip then being washed into a PCR tube for colony PCR analysis. Typically up to 20 colonies would be picked in this manner, if that many were present. The colony PCR was comprised of 15µl Red Taq (Sigma Aldrich), 1.5µl of each T7 primer (diluted ten-fold), 11µl Milli Q water, and the tip would be vigorously washed into this. T7 primers were chosen as they are present on the pET28a plasmid used for transformation, and allows for batches to be made at one point, allowing for multiple colony PCRs using different putative transformants to be tested simultaneously.

Colony PCRs were carried out slightly differently from previous PCRs:

1. Initial Denaturation Step: 98 ^{o}C for 2 minutes
2. Denaturation Cycle: 98 ^{o}C for 20 seconds
3. Combined Anneal: 50 ^{o}C for 30 seconds
4. Elongation Cycle: 72 ^{o}C for 105 seconds
5. Termination Step: 72 ^{o}C for 5 minutes

The above method applies to ADH D and E, but a variant CPCR was used for ADH F and G:

1. Initial Denaturation Step: 95 ^{o}C for 10 minutes
2. Denaturation Cycle: 95 ^{o}C for 30 seconds
3. Anneal Cycle: 50 ^{o}C for 45 seconds
4. Elongation Cycle: 72 ^{o}C for 60 seconds
5. Termination Step: 72 ^{o}C for 5 minutes

In both cases 40 cycles were run.

Post PCR the products were analysed by DNA gel electrophoresis in a similar manner to previous. Due to the large number of reactions typically run at this stage, larger gels were run, scaled up to 100ml, with 1wt% agarose (1g) and 5µl Sybrsafe. Visualisation is again via UV light.

Should positive results be noted in the DNA gel then the relevant colony can be picked from the secondary plates and transferred to liquid culture. At least two 10ml cultures (LB + kanamycin media) were prepared per transformant, and incubated overnight. The next day, two stocks can be prepared from one of the cultures (20vol% glycerol), and the remainder of the same culture can be miniprepped as per the Wizard Quick Protocol for MiniPreps DNA by centrifugation.

In order to check whether the transformant was as desired, the extracted plasmid was sent for sequencing to Eurofins and the plasmid was diluted as they recommended. Sequencing was carried out using T7 primers, which flank the multiple cloning site in pET28a.

Finally, if the sequencing was accurate, then the plasmid could be transformed into an expression strain of *E. coli*, either BL21 (DE3) or BL21 (DE3) pLysS. In this case transformation was by heatshock, rather than electroporation. 1µl of DNA was added to an aliquot (approximately 20µl) of cells and left on ice for 20 minutes. The aliquot was then incubated at 42 ^{o}C for 45 seconds before being returned to ice for two minutes. 1ml of LB media was then added and then incubated for one hour at 37 ^{o}C. After the recovery period 50µl-100µl was then plated out, this always ensured a very healthy number of colonies from which glycerol stocks could be obtained.

5.2. MICROBIOLOGY

All *E. coli* strains were grown using Lysogeny Broth (LB) with appropriate antibiotics (kanamycin 50µg ml^{-1} and chloramphenicol 34µg ml^{-1}) in the first instance. After inoculating with 1vol% of an overnight culture, 2L flasks containing 500ml LB would be grown until an optical density (OD) of approximately 0.8-1.2, after which IPTG was added to a final concentration of 1mM.

Flasks were then incubated for an additional 4.5 hours, continuing at 37 ^{o}C and 225rpm shaking. After this, the cells were centrifuged in 250ml portions using a Beckman-Coulter Allegra 25R centrifuge at 6000rpm for 10 minutes. The cells were then resuspended in 50ml of the original medium; typically resuspensions from two 250ml portions were combined into one 50ml Falcon tube. These were recentrifuged for 10 minutes at 4000 rpm using an Eppendorf 5810R centrifuge. The supernatant was discarded and the new pellets frozen at -20 ^{o}C until required. All centrifugation was carried out at 4 ^{o}C, and the preparation resulted in one cell pellet per 500ml original culture.

5.3. ENZYMOLOGY

The methods and provided data comply wherever possible with the recommendations of the STRENDA Consortium [Standards for Reporting Enzyme Data] (Kettner, 2007; Tipton et al., 2014), which aim to harmonise reported data amongst different research groups and journals to allow for easier reproducibility of data and for greater international integration of data.

5.3.1. EXTRACTION AND PURIFICATION

To prepare a sample of enzyme for study, a cell pellet was defrosted in 5ml of a buffer of 20mM imidazole, 300mM sodium chloride and 50mM Tris. A Complete Tablet Mini EDTA free Mini protease inhibitor (Roche) was also added, one tablet per 5ml pellet volume. After the pellet was entirely thawed, one of two methods would be used, sonication or freeze-thaw. In

the sonication variant the pellet was sonicated for 5 cycles of 15 seconds on, 30 seconds off at a wavelength of 21 microns on ice. This was sufficient to lyse the cells without excessive degradation to the enzyme. The freeze thaw method consisted of five cycles of being frozen at -80 ^{o}C, then being thawed at 42 ^{o}C in a water bath. Typically each phase lasted five minutes.

In addition, DNAse I (50U) was added after using the freeze thaw method, to reduce the viscosity of the lysate prior to further processing. Where the production strain did not already natively produce lysozyme, approximately 100,000 units were added at the same time as the DNase I.

To allow for the DNAse to work sufficiently, the pellet was shaken at 200rpm and 37 ^{o}C for circa 30 minutes to allow for sufficient mixing. The supplied buffer was also necessary for efficient DNA degradation, but instead of using the recommended volumes, an equal volume of the supplied buffer was added to the cell pellet as the DNAse itself, i.e. 50µl of both were added.

Irrespective of the overall method, freeze thaw or sonication, the pellet was then transferred into Eppendorfs and centrifuged at 13000rpm (~16000g) for 10 minutes to separate the soluble and insoluble fractions. The insoluble fraction was reserved, whilst the soluble fraction was filtered through 0.45µm and 0.22µm syringe filters in preparation for purification.

Purification was carried out by Immobilised Metal Affinity Chromatography (IMAC). All ADHs were expressed in pET28a, and have a hexa-histidine tag appended to the N terminal end of the target protein. Therefore, they were purified using a gradient of imidazole (20mM-1000mM) over a nickel-based column. The gravity column was prepared directly from His-bind resin (Novagen), of which 1ml were added. After washing with 5 column volumes (CV) of distilled water, 5CV of 400mM nickel sulphate ($NiSO_4 \cdot 6H_2O$) was added. Excess nickel sulphate was removed with 5CV water and 5CV of "His-Bind Buffer".

The column was stored in 20% ethanol after use, and equilibrated with His-Bind Buffer prior to reuse. This buffer comprised 20mM imidazole, 300mM sodium chloride and 50mM Tris. Various elution fractions were comprised using this buffer and the "His-Elute buffer", containing 1M imidazole rather than 20mM. To prepare the aforementioned fractions, these two buffers were blended as necessary, and each blend was named referring to the proportion of His-Elute Buffer, e.g. 20% His-Elute.

The initial filtered soluble fraction was loaded with a small quantity (typically ~1CV) of 0% His Elute to dilute it to roughly 5ml in volume and to ensure that the column never dried out post equilibration. After the flow through was collected, it was reloaded to the column, and the 2nd Flow through was then collected. Subsequent fractions were generated by adding 5ml of the relevant buffer blend to the column and collecting the run off.

Whilst purification itself was carried out at room temperature, all centrifugation and incubation steps were conducted at 4 ^{o}C, using ice where necessary.

Protein concentrations were determined by Bradford Assay, utilising a commercial reagent (Biorad) and bovine serum albumin as standard. Purity of proteins was determined by SDS-PAGE, using handcast 10% Bis-Tris gels composed as follows:

Component	Volume (separating) [ml]	Volume (stacking) [ml]
3.5x Buffer	2.84	1.00
30% Acrylamide	3.34	0.46
Milli-Q water	3.82	2.04
10wt% APS	0.05	0.04
TEMED	0.014	0.02

Where the 3.5x buffer contained 52.32g Bis-Tris per 200ml, at a pH of 6.5-6.8.

No further purification or isolation was conducted after fractions from the nickel column were collected. The fractions were stored at 4 ^{o}C as eluted and were not processed further. Assays were then conducted using raw fractions or serially diluted fractions as necessary.

5.3.2. STANDARD ALCOHOL DEHYROGENASE ASSAYS

Two methods were largely used, with the original method being used in Chapter 2, whilst an amended version was used in Chapter 3. The assay method was amended by necessity and by the need to ensure greater control over the temperature of the assay and improved solvent and chemical resistance.

5.3.2.1. INITIAL ASSAY METHOD

Assays were conducted at 50 ^{o}C using disposable cuvettes, either polystyrene or poly (methyl methacrylate) (PMMA), which have limited solvent resistance. Spectrophotometric data were acquired over approximately five minutes using a Cary 300 spectrophotometer set to 340nm. Typically, assays were carried out at pH 6.0, using 50mM citric acid buffer adjusted with sodium hydroxide and acetic acid as needed.

5.3.2.2. AMENDED ASSAY METHOD

Assays were conducted at 50 ^{o}C, but at pH 7.0 using 50mM sodium phosphate buffer adjusted with sodium hydroxide and hydrochloric acid as needed. Glass or quartz cuvettes were used, being washed with three volumes of 70% ethanol and dried in a static incubator at 50 ^{o}C prior to use. The final change is that a Cary 50 spectrophotometer was used to acquire the data.

5.3.2.3. COMMON ELEMENTS OF ASSAY METHOD

Spectrophotometric data were acquired over a timeframe of approximately five minutes including equilibration time. Rates were taken from the initial, linear, portion of all traces. Thus data were from direct continuous assays, most usually over a period of thirty seconds or more. Where it was not possible to have such a lengthy linear trace, rates were still taken from the most appropriately long and linear portion of the trace at the initial phase of the assay, excluding any obvious discrepancy. Generally all given data are the results of triplicate assays run identically, with the exception of substrate screening. In this case, three different assays were run, varying the enzyme concentration with each substrate, and if enzyme proportionality was maintained, then the specific activity given is an average calculated from

all three assays. If proportionality was not maintained using the highest concentration of enzyme, then the highest value was discarded and the average of the two appropriate values given instead.

In a typical assay reagents were added sequentially, with the buffer as the largest volume being added first. The buffer was usually the only reagent to be preincubated, and it was kept at 70 ^{o}C in a water bath prior to addition. This enabled a relatively rapid acclimatisation to assay conditions. After the buffer, any solvent was added, then substrate and cofactor. The cuvette was then removed and shaken, if the assay temperature was 50 ^{o}C or below; above this, the mixing was done *in situ,* along with the addition of the reagents. After this, the enzyme was added to start the reaction, and a further mixing step was done as previously. Care was also taken to ensure a minimum of bubbles was created, by lightly tapping the base of the cuvette immediately prior to inserting it for reading. All assays comprised a total volume of 1ml, with varying volumes of enzyme, substrate and cofactor as necessary.

Controls were done by varying the order of introducing components to the assay, for example to rule out the possibility of aldehydes reacting with the enzymes being tested, they were preincubated with the substrate and the assay started with the cofactor. Another notable control was a no-substrate control, to rule out incidental effects upon the NAD(P)H readings.

5.3.3. STABILITY ANALYSIS

All forms of stability analysis used the same assays as given above. When testing thermostability, enzyme was preincubated at various temperatures then put on ice. Standard assay techniques were then used.

Optimal temperature work used standard assays with the exception of the assay temperature. To accommodate higher temperatures, mixing was achieved *in situ* to avoid excessive heat loss and to prevent injury. The cuvette was also capped to prevent evaporation of substrate, products or other assay components.

Work with solvents used standard assay techniques, with solvents being directly added to the cuvette immediately prior to the assay.

5.3.4. THERMAL SHIFT ASSAYS

The dye SYPRO®-Orange (Invitrogen, Life Technologies, UK) was added at a final dilution of 90-fold and enzymes were added at an approximate concentration of 0.1-1.0 mg ml^{-1} to optically clear Genie® II tubes to a final volume of 25µl. Samples were heated from 25-105 °C at a rate of 0.05 °C s^{-1}, with fluorescence measured at 470nm excitation and 555nm emission in a Genie® II real-time fluorescence detection instrument (Optigene, UK). The instrument and instruction was kindly provided by Dr Nick Morant, GeneSys Biotech Ltd.

The derivative of the fluorescence was recorded and plotted by the instrument, and the peak recorded to provide an estimate of T_m.

5.4. ANALYTICAL CHEMISTRY

For chemical analysis, the biocatalytic reaction was carried out in 15ml centrifuge tubes, using identical assay components. Thus the scale of reaction was 1ml. The reaction was carried out either for one hour or 16 hours whilst capped, at 50 ^{o}C and vigorously shaken.

Extraction was by one reaction volume (1ml) of ethyl acetate. The solvent was vigorously shaken to ensure efficient mixing, and then briefly centrifuged at 4 ^{o}C to ensure separation of solvent and water. The solvent was either directly analysed by GC/MS or diluted into fresh ethyl acetate as necessary for GC/MS analysis.

A standard GC/MS method was used, ramping from 40 ^{o}C to 250 ^{o}C over ten minutes; the column used was a 30m DB-FFAP (Agilent Technologies), on an Agilent Technologies 7890B GC/MS system.

In all cases external standards were used, five points ranging from 2mM through to 10mM.

6. REFERENCES

Abraham, M.A., Nguyen, N., 2003. “Green Engineering: Defining the Principles” - Results from the Sandestin Conference. Environ. Prog. 22, 233–236. doi:10.1002/ep.670220410

Alves, L.A., Almeide e Silva, J.B., Giulietti, M., 2007. Solubility of D -Glucose in Water and Ethanol / Water Mixtures. J. Chem. Eng. Data 52, 2166–2170.

Anastas, P., Eghbali, N., 2010. Green chemistry: principles and practice. Chem. Soc. Rev. 39, 301–312. doi:10.1039/b918763b

Anastas, P.T., Kirchhoff, M.M., 2002. Origins, current status, and future challenges of green chemistry. Acc. Chem. Res. 35, 686–94.

Anastas, P.T., Zimmerman, J.B., 2003. Through the 12 Principles of Green Engineering. Environ. Sci. Technol. 95. doi:10.1039/b411954c

Anthony, J.R., Anthony, L.C., Nowroozi, F., Kwon, G., Newman, J.D., Keasling, J.D., 2009. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. Metab. Eng. 11, 13–9. doi:10.1016/j.ymben.2008.07.007

Assareh, R., Shahbani Zahiri, H., Akbari Noghabi, K., Aminzadeh, S., Bakhshi Khaniki, G., 2012. Characterization of the newly isolated Geobacillus sp. T1, the efficient cellulase-producer on untreated barley and wheat straws. Bioresour. Technol. 120, 99–105. doi:10.1016/j.biortech.2012.06.027

Atsumi, S., Liao, J.C., 2008. Metabolic engineering for advanced biofuels production from Escherichia coli. Curr. Opin. Biotechnol. 19, 414–9. doi:10.1016/j.copbio.2008.08.008

Atsumi, S., Wu, T.-Y., Eckl, E.-M., Hawkins, S.D., Buelter, T., Liao, J.C., 2010. Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl. Microbiol. Biotechnol. 85, 651–7. doi:10.1007/s00253-009-2085-6

Azadi, P., Carrasquillo-Flores, R., Pagán-Torres, Y.J., Gürbüz, E.I., Farnood, R., Dumesic, J.A., 2012. Catalytic conversion of biomass using solvents derived from lignin. Green Chem. 14, 1573. doi:10.1039/c2gc35203f

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75. doi:10.1186/1471-2164-9-75

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50, 4402–10. doi:10.1021/bi2002289

Barnard, D., Casanueva, A., Tuffin, M.I., Cowan, D., 2010. Extremophiles in biofuel synthesis. Environ. Technol. 31, 871–888. doi:10.1080/09593331003710236

Bergquist, P.L., Morgan, H.W., Saul, D., 2014. Selected Enzymes from Extreme Thermophiles with Applications in Biotechnology. Curr. Biotechnol. 3, 45–59.

Bilitewski, B., 2012. The Circular Economy and its Risks. Waste Manag. 32, 1–2.

doi:10.1016/j.wasman.2011.10.004

Blacker, J., Williams, M.T., 2011. Future Trends and Challenges, in: Blacker, A.J., Williams, M.T. (Eds.), Pharmaceutical Process Development: Current Chemical and Engineering Challenges. Royal Society of Chemistry.

Blikstad, C., Widersten, M., 2010. Functional characterization of a stereospecific diol dehydrogenase, FucO, from Escherichia coli: Substrate specificity, pH dependence, kinetic isotope effects and influence of solvent viscosity. J. Mol. Catal. B Enzym. 66, 148–155. doi:10.1016/j.molcatb.2010.04.010

Bozell, J.J., Petersen, G.R., 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. Green Chem. 12, 539. doi:10.1039/b922014c

Brenner, K., You, L., Arnold, F.H., 2008. Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol. 26, 483–9. doi:10.1016/j.tibtech.2008.05.004

Burbelo, P.D., Ching, K.H., Han, B.L., Klimavicz, C.M., Iadarola, M.J., 2010. Synthetic biology for translational research. Am. J. Transl. Res. 2, 381–9.

Carothers, J.M., Goler, J.A., Keasling, J.D., 2009. Chemical synthesis using synthetic biology. Curr. Opin. Biotechnol. 20, 498–503. doi:10.1016/j.copbio.2009.08.001

Carvalho, C.C.C.R. De, Fonseca, M.M.R., 2002. Maintenance of cell viability in the biotransformation of (−) -carveol with whole cells of Rhodococcus erythropolis. J. Mol. Catal. B 19-20, 389–398.

Ceccarelli, C., Liang, Z.-X., Strickler, M., Prehna, G., Goldstein, B.M., Klinman, J.P., Bahnson, B.J., 2004. Crystal structure and amide H/D exchange of binary complexes of alcohol dehydrogenase from Bacillus stearothermophilus: insight into thermostability and cofactor binding. Biochemistry 43, 5266–77. doi:10.1021/bi049736p

Cha, M., Chambliss, G.H., 2013. Cloning and sequence analysis of the heat-stable acrylamidase from a newly isolated thermophilic bacterium, Geobacillus thermoglucosidasius AUT-01. Biodegradation 24, 57–67. doi:10.1007/s10532-012-9557-6

Chang, T., Yao, S., 2011. Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives. Appl. Microbiol. Biotechnol. 13–27. doi:10.1007/s00253-011-3456-3

Chenault, H.K., Whitesides, G.M., 1987. Regeneration of nicotinamide cofactors for use in organic synthesis. Appl. Biochem. Biotechnol. 14, 147–97.

Chiyanzu, I., Cowan, D., Burton, S.G., 2010. Immobilization of Geobacillus pallidus RAPc8 nitrile hydratase (NHase) reduces substrate inhibition and enhances thermostability. J. Mol. Catal. B Enzym. 63, 109–115. doi:10.1016/j.molcatb.2009.12.011

Clark, J., 2007. Green chemistry for the second generation biorefinery—sustainable chemical manufacturing based on biomass. J. Chem. Technol. Biotechnol. 609, 603–609. doi:10.1002/jctb

Clark, J.H., Budarin, V., Deswarte, F.E.I., Hardy, J.J.E., Kerton, F.M., Hunt, A.J., Luque, R., Macquarrie, D.J., Milkowski, K., Rodriguez, A., Samuel, O., Tavener, S.J., White, R.J., Wilson, A.J., 2006. Green chemistry and the biorefinery: a partnership for a sustainable future. Green Chem. 8, 853. doi:10.1039/b604483m

Clouthier, C.M., Pelletier, J.N., 2012. Expanding the organic toolbox: a guide to integrating

biocatalysis in synthesis. Chem. Soc. Rev. 41, 1585–1605. doi:10.1039/c2cs15286j

Commichau, F.M., Pietack, N., Stülke, J., 2013. Essential genes in Bacillus subtilis: a re-evaluation after ten years. Mol. Biosyst. 9, 1068–75. doi:10.1039/c3mb25595f

Connor, M.R., Atsumi, S., 2010. Synthetic biology guides biofuel production. J. Biomed. Biotechnol. doi:10.1155/2010/541698

Conway, T., Ingram, L.O., 1989. Similarity of Escherichia coli propanediol oxidoreductase (fucO product) and an ununsual alcohol dehydrogenase from Zymononas mobilis and Saccharomyces cerevisiae. J. Bacteriol. 171, 3754.

Cowan, D., 1997. Thermophilic proteins: Stability and function in aqueous and organic solvents. Comp. Biochem. Physiol. Part A Physiol. 118, 429–438. doi:10.1016/S0300-9629(97)00004-2

Cripps, R.E., Eley, K.L., Leak, D.J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., Atkinson, T., 2009. Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production. Metab. Eng. 11, 398–408. doi:10.1016/j.ymben.2009.08.005

Dale, D.J., 2011. The Design of Safe Chemical Reactions : It ' s No Accident, in: Blacker, A.J., Williams, M.T. (Eds.), Pharmaceutical Process Development: Current Chemical and Engineering Challenges. Royal Society of Chemistry, pp. 160–177.

Dautzenberg, G., Gerhardt, M., Kamm, B., 2011. Bio based fuels and fuel additives from lignocellulose feedstock via the production of levulinic acid and furfural. Holzforschung 65, 439–451. doi:10.1515/HF.2011.081

de Carvalho, C.C.C.R., 2011. Enzymatic and whole cell catalysis: finding new strategies for old processes. Biotechnol. Adv. 29, 75–83. doi:10.1016/j.biotechadv.2010.09.001

De Man, R., Friege, H., 2016. Circular economy : european policy on shaky ground. Waste Manag. Res. 34, 93–95. doi:10.1177/0734242X15626015

Dellomonaco, C., Fava, F., Gonzalez, R., 2010. The path to next generation biofuels: successes and challenges in the era of synthetic biology. Microb. Cell Fact. 9, 3. doi:10.1186/1475-2859-9-3

Desai, A.A., 2011. Sitagliptin manufacture: a compelling tale of green chemistry, process intensification, and industrial asymmetric catalysis. Angew. Chem. Int. Ed. Engl. 50, 1974–6. doi:10.1002/anie.201007051

Di Luccio, E., Elling, R.A., Wilson, D.K., 2006. Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. Biochem. J. 400, 105–14. doi:10.1042/BJ20060660

Dordick, J., 1989. Enzymatic catalysis in monophasic organic solvents. Enzyme Microb. Technol. 11, 194–211.

Doukyu, N., Ogino, H., 2010. Organic solvent-tolerant enzymes. Biochem. Eng. J. 48, 270–282. doi:10.1016/j.bej.2009.09.009

Drury, J., Penning, T.M., n.d. Aldo-Keto Reductase Superfamily Database [WWW Document]. URL https://www.med.upenn.edu/akr/ (accessed 9.28.15).

Du, J., Shao, Z., Zhao, H., 2011. Engineering microbial factories for synthesis of value-added products. J. Ind. Microbiol. Biotechnol. 38, 873–90. doi:10.1007/s10295-011-0970-3

Dunn, P.J., 2011. The importance of green chemistry in process research and development., in: Blacker, A.J., Williams, M.T. (Eds.), RSC Drug Discovery Series No.9 Pharmaceutical Process Development: Current Chemical and Engineering Challenges. Royal Society of Chemistry, pp. 117–137. doi:10.1039/c1cs15041c

Edegger, K., Mang, H., Faber, K., Gross, J., Kroutil, W., 2006. Biocatalytic oxidation of sec-alcohols via hydrogen transfer. J. Mol. Catal. A Chem. 251, 66–70. doi:10.1016/j.molcata.2006.02.007

Ehrensberger, A.H., Wilson, D.K., 2004. Structural and Catalytic Diversity in the Two Family 11 Aldo-keto Reductases. J. Mol. Biol. 337, 661–673. doi:10.1016/j.jmb.2004.01.059

Eisenthal, R., Danson, M.J., Hough, D.W., 2007. Catalytic efficiency and kcat/KM: a useful comparator? Trends Biotechnol. 25, 247–9. doi:10.1016/j.tibtech.2007.03.010

Ellen MacArthur Foundation, 2015. Towards a circular economy: Business rationale for an accelerated transitions [WWW Document]. URL http://www.ellenmacarthurfoundation.org/publications/towards-a-circular-economy-business-rationale-for-an-accelerated-transition (accessed 8.16.16).

Elleuche, S., Fodor, K., Klippel, B., von der Heyde, A., Wilmanns, M., Antranikian, G., 2013. Structural and biochemical characterisation of a NAD(+)-dependent alcohol dehydrogenase from Oenococcus oeni as a new model molecule for industrial biotechnology applications. Appl. Microbiol. Biotechnol. 8963–8975. doi:10.1007/s00253-013-4725-0

Elleuche, S., Schäfers, C., Blank, S., Schröder, C., Antranikian, G., 2015. Exploration of extremophiles for high temperature biotechnological processes. Curr. Opin. Microbiol. 25, 113–119. doi:10.1016/j.mib.2015.05.011

Elleuche, S., Schröder, C., Sahm, K., Antranikian, G., 2014. Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. Curr. Opin. Biotechnol. 29, 116–123. doi:10.1016/j.copbio.2014.04.003

Erickson, B., Nelson, J.E., Winters, P., 2012. Perspective on opportunities in industrial biotechnology in renewable chemicals. Biotechnol. J. 7, 176–85. doi:10.1002/biot.201100069

Extance, J., Crennell, S.J., Eley, K., Cripps, R., Hough, D.W., Danson, M.J., 2013. Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in Geobacillus thermoglucosidasius. Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 2104–2115. doi:10.1107/S0907444913020349

Extance, J.P., 2012. Bioethanol Production : Characterisation of a Bifunctional Alcohol Dehydrogenase from Geobacillus thermoglucosidasius. PhD Thesis. University of Bath.

Felnagle, E.A., Chaubey, A., Noey, E.L., Houk, K.N., Liao, J.C., 2012. Engineering synthetic recursive pathways to generate non-natural small molecules. Nat. Chem. Biol. 8, 518–526. doi:10.1038/nchembio.959

Fischbach, M., Voigt, C.A., 2010. Prokaryotic gene clusters: a rich toolbox for synthetic biology. Biotechnol. J. 5, 1277–96. doi:10.1002/biot.201000181

Fong, J.C.N., Svenson, C.J., Nakasugi, K., Leong, C.T.C., Bowman, J.P., Chen, B., Glenn, D.R., Neilan, B.A., Rogers, P.L., 2006. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. Extremophiles 10, 363–72. doi:10.1007/s00792-006-0507-2

Frock, A.D., Kelly, R.M., 2012. Extreme thermophiles: moving beyond single-enzyme biocatalysis. Curr. Opin. Chem. Eng. 1, 363–372. doi:10.1016/j.coche.2012.07.003

Gao, H., Zhuo, Y., Ashforth, E., Zhang, L., 2010. Engineering of a genome-reduced host: practical application of synthetic biology in the overproduction of desired secondary metabolites. Protein Cell 1, 621–6. doi:10.1007/s13238-010-0073-3

Gao, X., Ni, K., Zhao, C., Ren, Y., Wei, D., 2014. Enhancement of the activity of enzyme immobilized on polydopamine-coated iron oxide nanoparticles by rational orientation of formate dehydrogenase. J. Biotechnol. 188, 36–41. doi:10.1016/j.jbiotec.2014.07.443

Garner, A., Keoleian, G.A., 1995. Industrial Ecology : An Introduction, in: Pollution Prevention Educational Resource Compendium. National Pollution Prevention Center for Higher Education, Ann Arbor, pp. 1–32.

Geboers, J.A., Vyver, S. Van De, Ooms, R., Beeck, B. Op de, Jacobs, P.A., Sels, B.F., 2011. Chemocatalytic conversion of cellulose: opportunities, advances and pitfalls. Catal. Sci. Technol. 1, 714–726. doi:10.1039/c1cy00093d

Gekko, K., Timasheff, S.N., 1981. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. Biochemistry 20, 4667–4676. doi:10.1021/bi00519a023

Girilal, M., Fayaz, A.M., Balaji, P.M., Kalaichelvan, P.T., 2013. Augmentation of PCR efficiency using highly thermostable gold nanoparticles synthesized from a thermophilic bacterium, Geobacillus stearothermophilus. Colloids Surf. B. Biointerfaces 106, 165–9. doi:10.1016/j.colsurfb.2012.12.038

Goihberg, E., Dym, O., Tel-Or, S., Shimon, L., Frolow, F., Peretz, M., Burstein, Y., 2008. Thermal stabilization of the protozoan Entamoeba histolytica alcohol dehydrogenase by a single proline substitution. Proteins Struct. Funct. Genet. 72, 711–719. doi:10.1002/prot.21946

Goldberg, K., Krueger, A., Meinhardt, T., Kroutil, W., Mautner, B., Liese, A., 2008. Novel immobilization routes for the covalent binding of an alcohol dehydrogenase from Rhodococcus ruber DSM 44541. Tetrahedron: Asymmetry 19, 1171–1173. doi:10.1016/j.tetasy.2008.04.034

Goldberg, K., Schroer, K., Lütz, S., Liese, A., 2007a. Biocatalytic ketone reduction--a powerful tool for the production of chiral alcohols--part I: processes with isolated enzymes. Appl. Microbiol. Biotechnol. 76, 237–48. doi:10.1007/s00253-007-1002-0

Goldberg, K., Schroer, K., Lütz, S., Liese, A., 2007b. Biocatalytic ketone reduction--a powerful tool for the production of chiral alcohols-part II: whole-cell reductions. Appl. Microbiol. Biotechnol. 76, 249–55. doi:10.1007/s00253-007-1005-x

Grimaldi, J., Collins, C.H., Belfort, G., 2016. Towards cell-free isobutanol production: Development of a novel immobilized enzyme system. Biotechnol. Prog. 32, 66–73. doi:10.1002/btpr.2197

Gu, Y., Jérôme, F., 2010. Glycerol as a sustainable solvent for green chemistry. Green Chem. 12, 1127. doi:10.1039/c001628d

Guagliardi, A., Martino, M., Iaccarino, I., De Rosa, M., Rossi, M., Bartolucci, S., 1996. Purification and characterization of the alcohol dehydrogenase from a novel strain of Bacillus stearothermophilus growing at 70 degrees C. Int. J. Biochem. Cell Biol. 28, 239–46.

Gulbis, J.M., Mann, S., MacKinnon, R., 1999. Structure of a Voltage-Dependent K+ Channel β Subunit. Cell 97, 943–952. doi:10.1016/S0092-8674(00)80805-3

Guterl, J.-K., Sieber, V., 2013. Biosynthesis "debugged": Novel bioproduction strategies. Eng. Life Sci. 13, 4–18. doi:10.1002/elsc.201100231

Guy, J.E., Isupov, M.N., Littlechild, J.A., 2003. The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon Aeropyrum pernix. J. Mol. Biol. 331, 1041–1051. doi:10.1016/S0022-2836(03)00857-X

Haki, G., 2003. Developments in industrially important thermostable enzymes: a review. Bioresour. Technol. 89, 17–34. doi:10.1016/S0960-8524(03)00033-6

Hale, V., Keasling, J.D., Renninger, N.S., Diagana, T.T., 2007. Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. Am. J. Trop. Med. Hyg. 77, 198–202.

Hall, M., Bommarius, A.S., 2011. Enantioenriched Compounds via Enzyme-Catalyzed Redox Reactions. Chem. Rev. 4088–4110. doi:dx.doi.org/10.1021/cr200013n

Hatti-Kaul, R., Törnvall, U., Gustafsson, L., Börjesson, P., 2007. Industrial biotechnology for the production of bio-based chemicals--a cradle-to-grave perspective. Trends Biotechnol. 25, 119–24. doi:10.1016/j.tibtech.2007.01.001

Hawwa, R., Aikens, J., Turner, R.J., Santarsiero, B.D., Mesecar, A.D., 2009. Structural basis for thermostability revealed through the identification and characterization of a highly thermostable phosphotriesterase-like lactonase from Geobacillus stearothermophilus. Arch. Biochem. Biophys. 488, 109–120. doi:10.1016/j.abb.2009.06.005

Hedlund, J., Jörnvall, H., Persson, B., 2010. Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement. BMC Bioinformatics 11, 534. doi:10.1186/1471-2105-11-534

Hilterhaus, L., Liese, A., 2009. Applications of Reaction Engineering to Industrial Biotransformations, in: Tao, J., Lin, G.-Q., L, A. (Eds.), Biocatalysis for the Pharmaceutical Industry: Discovery, Development and Manufacturing. John Wiley & Sons Asia (Pte) Ltd.

Hollmann, F., Arends, I.W.C.E., Buehler, K., Schallmey, A., Bühler, B., 2011a. Enzyme-mediated oxidations for the chemist. Green Chem. 13, 226. doi:10.1039/c0gc00595a

Hollmann, F., Arends, I.W.C.E., Holtmann, D., 2011b. Enzymatic reductions for the chemist. Green Chem. 13, 2285–2313. doi:10.1039/c1gc15424a

Höllrigl, V., Hollmann, F., Kleeb, A.C., Buehler, K., Schmid, A., 2008. TADH, the thermostable alcohol dehydrogenase from Thermus sp. ATN1: a versatile new biocatalyst for organic synthesis. Appl. Microbiol. Biotechnol. 81, 263–73. doi:10.1007/s00253-008-1606-z

Honda, K., Hara, N., Cheng, M., Nakamura, A., Mandai, K., Okano, K., Ohtake, H., 2016. In vitro metabolic engineering for the salvage synthesis of NAD+. Metab. Eng. 35, 114–120. doi:10.1016/j.ymben.2016.02.005

Huisman, G.W., Liang, J., Krebber, A., 2010. Practical chiral alcohol manufacture using ketoreductases. Curr. Opin. Chem. Biol. 14, 122–9. doi:10.1016/j.cbpa.2009.12.003

Hussein, A.H., Lisowska, B.K., Leak, D.J., 2015. The Genus Geobacillus and Their Biotechnological Potential, in: Sariaslani, S., Gadd, G.M. (Eds.), Advances in Applied Microbiology. pp. 1–48.

Hutchins, L.M., Hunter, L., Ehya, N., Gibbs, M.D., Bergquist, P.L., Hutton, C.A., 2004. Highly enantioselective recombinant thermoalkalophilic lipases from Geobacillus and Bacillus sp. Tetrahedron: Asymmetry 15, 2975–2980. doi:10.1016/j.tetasy.2004.07.041

Ismaiel, A.A., Zhu, C.X., Colby, G.D., Chen, J.S., 1993. Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of Clostridium beijerinckii. J. Bacteriol. 175, 5097–105.

Jacquet, N., Haubruge, E., Richel, A., 2015. Production of biofuels and biomolecules in the framework of circular economy: A regional case study. Waste Manag. Res. J. Int. Solid Wastes Public Clean. Assoc. 12, 1121–1126 NV – 33. doi:10.1177/0734242X15613154

Jarboe, L.R., 2011. YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. Appl. Microbiol. Biotechnol. 89, 249–57. doi:10.1007/s00253-010-2912-9

Jarboe, L.R., Zhang, X., Wang, X., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2010. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. J. Biomed. Biotechnol. 761042. doi:10.1155/2010/761042

Jeon, Y.J., Fong, J.C.N., Riyanti, E.I., Neilan, B.A., Rogers, P.L., Svenson, C.J., 2008. Heterologous expression of the alcohol dehydrogenase (adhI) gene from Geobacillus thermoglucosidasius strain M10EXG. J. Biotechnol. 135, 127–33. doi:10.1016/j.jbiotec.2008.02.018

Ji, X.-J., Huang, H., Ouyang, P.-K., 2011. Microbial 2,3-butanediol production: a state-of-the-art review. Biotechnol. Adv. 29, 351–64. doi:10.1016/j.biotechadv.2011.01.007

Ji, Y., Mao, G., Wang, Y., Bartlam, M., 2013. Crystallization and preliminary X-ray characterization of an NAD(P)-dependent butanol dehydrogenase A from Geobacillus thermodenitrificans NG80-2. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 69, 184–7. doi:10.1107/S1744309113000766

Jiménez-González, C., Constable, D.J.C., Ponder, C.S., 2012. Evaluating the “greenness” of chemical processes and products in the pharmaceutical industry--a green metrics primer. Chem. Soc. Rev. 41, 1485–98. doi:10.1039/c1cs15215g

Johnson, D.T., Taconi, K.A., 2007. The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. Environ. Prog. 26, 338–348. doi:10.1002/ep

Jörnvall, H., Hedlund, J., Bergman, T., Kallberg, Y., Cederlund, E., Persson, B., 2013. Origin and evolution of medium chain alcohol dehydrogenases. Chem. Biol. Interact. 202, 91–6. doi:10.1016/j.cbi.2012.11.008

Joshi, B., Bhatt, M.R., Sharma, D., Joshi, J., Malla, R., Sreerama, L., 2011. Lignocellulosic ethanol production: current practices and recent developments. Biotechnol. Mol. Biol. Rev. 6, 172–182.

Kallberg, Y., Oppermann, U., Persson, B., 2010. Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. FEBS J. 277, 2375–2386. doi:10.1111/j.1742-4658.2010.07656.x

Kamm, B., Kamm, M., 2007. Biorefineries – Multi Product Processes. Adv Biochem Engin/Biotechnol 105, 175–204. doi:10.1007/10

Kaul, P., Banerjee, A., Banerjee, U., 2004. Opportunities for the Pharmaceutical Industry: Key

Biotransformation Technologies of the Future. Drug Discov. World 80–86.

Keasling, J.D., 2008. Synthetic biology for synthetic chemistry. ACS Chem. Biol. 3, 64–76. doi:10.1021/cb7002434

Kettner, C., 2007. Good Publication Practice As A Prerequisite for Comparable Enzyme Data ? In Silico Biol. 7, 57–64.

Kiser, B., 2016. Getting the circulation going. Nature 531, 443–446.

Knezevic-Jugovic, Z.D., Bezbradica, D.I., Mijin, D.Z., Antov, M.G., 2011. Introduction to the Field of Enzyme Stabilization and Immobilization, in: Minteer, S.D. (Ed.), Enzyme Stabilization and Immobilization: Methods and Protocols, Methods in Molecular Biology 679. Humana Press, Totowa, NJ, p. 230. doi:10.1007/978-1-60761-895-9

Knoll, M., Pleiss, J., 2008. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence–structure–function relationship. Protein Sci. 1689–1697. doi:10.1110/ps.035428.108.Protein

Kolb, H.C., Finn, M.G., Sharpless, K.B., 2001. Click chemistry Diverse chemical funtion from a few good reactions. Angew. Chem. Int. Ed. Engl. 40, 2004–2021.

Korkhin, Y., Kalb(Gilboa), A.J., Peretz, M., Bogin, O., Burstein, Y., Frolow, F., 1998. NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of Clostridium beijerinckii and Thermoanaerobacter brockii. J. Mol. Biol. 278, 967–981. doi:10.1006/jmbi.1998.1750

Kuiken, T., 2013. Converging technologies for a smarter health and wellness future, in: ICTs and the Health Sector. OECD Publishing, pp. 159–177. doi:10.1787/9789264202863-12-en

Kulig, J., Frese, A., Kroutil, W., Pohl, M., Rother, D., 2013. Biochemical characterization of an alcohol dehydrogenase from Ralstonia sp. Biotechnol. Bioeng. 110, 1838–1848. doi:10.1002/bit.24857

Kulishova, L., Dimoula, K., Jordan, M., Wirtz, A., Hofmann, D., Santiago-Schübel, B., Fitter, J., Pohl, M., Spiess, A.C., 2010. Factors influencing the operational stability of NADPH-dependent alcohol dehydrogenase and an NADH-dependent variant thereof in gas/solid reactors. J. Mol. Catal. B Enzym. 67, 271–283. doi:10.1016/j.molcatb.2010.09.005

Kumar, A., Shen, P.S., Descoteaux, S., Pohl, J., Bailey, G., Samuelson, J., 1992. Cloning and expression of an NADP(+)-dependent alcohol dehydrogenase gene of Entamoeba histolytica. Proc Natl Acad Sci U S A 89, 10188–10192.

Kung, Y., Runguphan, W., Keasling, J.D., 2012. From fields to fuels: recent advances in the microbial production of biofuels. ACS Synth. Biol. 1, 498–513. doi:10.1021/sb300074k

Kurbanoglu, E.B., Zilbeyaz, K., Kurbanoglu, N.I., 2011. Cryptococcus laurentii as a new biocatalyst for the asymmetric reduction of substituted acetophenones. Tetrahedron: Asymmetry 22, 345–350. doi:10.1016/j.tetasy.2011.01.014

Laird, D.W., Rowen, C.C., Machold, T., May, P.M., Hefter, G., 2013. Volatile Products from the Degradation of Organics in a Synthetic Bayer Liquor. Ind. Eng. Chem. Res. 52, 3613–3617. doi:10.1021/ie3024824

Laird, T., 2011. Process Research and Development in the Pharmaceutical Industry : Origins , Evolution and Progress, in: Blacker, A.J., Mike T Williams (Eds.), Pharmaceutical Process Development: Current Chemical and Engineering Challenges. Royal Society of Chemistry.

Lapthorn, A.J., Zhu, X., Ellis, E.M., 2013. The diversity of microbial aldo/keto reductases from Escherichia coli K12. Chem. Biol. Interact. 202, 168–177. doi:10.1016/j.cbi.2012.10.008

Lei, J., Zhou, Y.F., Li, L.F., Su, X.D., 2009. Structural and biochemical analyses of YvgN and YtbE from Bacillus subtilis. Protein Sci. 18, 1792–1800. doi:10.1002/pro.178

Lerchner, A., Jarasch, A., Meining, W., Schiefner, A., Skerra, A., 2013. Crystallographic analysis and structure-guided engineering of NADPH-dependent Ralstonia sp. alcohol dehydrogenase toward NADH cosubstrate specificity. Biotechnol. Bioeng. 110, 2803–14. doi:10.1002/bit.24956

Leuchs, S., Greiner, L., 2011. Alcohol Dehydrogenase from Lactobacillus brevis : A Versatile Robust Catalyst for Enantioselective Transformations. Chem Biochem Eng Q 25, 267–281.

Liang, J., Luo, Y., Zhao, H., 2011. Synthetic biology: putting synthesis into biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 3, 7–20. doi:10.1002/wsbm.104

Lin, J., Aoll, J., Niclass, Y., Velazco, M.I., Wünsche, L., Pika, J., Starkenmann, C., 2013. Qualitative and quantitative analysis of volatile constituents from latrines. Environ. Sci. Technol. 47, 7876–82. doi:10.1021/es401677q

Lin, P.P., Rabe, K.S., Takasumi, J.L., Kadisch, M., Arnold, F.H., Liao, J.C., 2014. Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius. Metab. Eng. 24, 1–8. doi:10.1016/j.ymben.2014.03.006

Littlechild, J., Novak, H., James, P., Sayer, C., 2013. Mechanisms of Thermal Stability Adopted by Thermophilic Proteins and Their Use in White Biotechnology, in: Thermophilic Microbes in Environmental and Industrial Biotechnology. Springer Netherlands, Dordrecht, pp. 481–507. doi:10.1007/978-94-007-5899-5_19

Littlechild, J.A., 2015. Enzymes from Extreme Environments and Their Industrial Applications. Front. Bioeng. Biotechnol. 3, 161. doi:10.3389/fbioe.2015.00161

Littlechild, J.A., 2011. Thermophilic archaeal enzymes and applications in biocatalysis. Biochem. Soc. Trans. 39, 155–158. doi:10.1042/BST0390155

Liu, X., Dong, Y., Zhang, J., Zhang, A., Wang, L., Feng, L., 2009. Two novel metal-independent long-chain alkyl alcohol dehydrogenases from Geobacillus thermodenitrificans NG80-2. Microbiology 155, 2078–85. doi:10.1099/mic.0.027201-0

Long, S.B., Campbell, E.B., MacKinnon, R., 2005. Crystal structure of a Mammalian Voltage-Dependent Shaker Family K+ Channel. Science 309, 897–903.

Lu, D., Williams, P.G., Wang, G., 2009. Metabolic Engineering for the Development and Manufacturing of Pharmaceuticals, in: Tao, J., Lin, G.-Q., L, A. (Eds.), Biocatalysis for the Pharmaceutical Industry: Discovery, Development and Manufacturing. John Wiley & Sons Asia (Pte) Ltd.

Lü, J., Sheahan, C., Fu, P., 2011. Metabolic engineering of algae for fourth generation biofuels production. Energy Environ. Sci. 4, 2451. doi:10.1039/c0ee00593b

Ma, C., Zhang, L., Dai, J., Xiu, Z., 2010. Relaxing the coenzyme specificity of 1,3-propanediol oxidoreductase from Klebsiella pneumoniae by rational design. J. Biotechnol. 146, 173–178. doi:10.1016/j.jbiotec.2010.02.005

Ma, K., Adams, M.W.W., 1999. Zinc-Containing Alcohol Dehydrogenase from the Hyperthermophilic Archaeon Pyrococcus furiosus An Unusual Oxygen-Sensitive , Iron- and Zinc-Containing Alcohol Dehydrogenase from the Hyperthermophilic Archaeon

Pyrococcus furiosus. J. Bacteriol. 181, 1163–1170.

Ma, S.K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G.W., Sheldon, R. a., 2010. A green-by-design biocatalytic process for atorvastatin intermediate. Green Chem. 12, 81. doi:10.1039/b919115c

Mackenzie, K., Eddy, C., Ingram, L., 1989. Modulation of alcohol dehydrogenase isoenzyme levels in Zymomonas mobilis by iron and zinc. J. Bacteriol. 171, 1063–1067.

Mączyski, A., Shaw, D.G., Goral, M., Winiewska-Gocłowska, B., 2008. IUPAC-NIST Solubility Data Series. 86. Ethers and Ketones with Water. Part 4. C4 and C5 Ketones with Water. J. Phys. Chem. Ref. Data 37, 1517. doi:10.1063/1.2945626

Man, H., Gargiulo, S., Frank, A., Hollmann, F., Grogan, G., 2014a. Structure of the NADH-dependent thermostable alcohol dehydrogenase TADH from Thermus sp. ATN1 provides a platform for engineering specificity and improved compatibility with inorganic cofactor-regeneration catalysts. J. Mol. Catal. B Enzym. 105, 1–6. doi:10.1016/j.molcatb.2014.03.013

Man, H., Kedziora, K., Kulig, J., Frank, A., Lavandera, I., Gotor-Fernandez, V., Rother, D., Hart, S., Turkenburg, J.P., Grogan, G., 2014b. Structures of alcohol dehydrogenases from Ralstonia and Sphingobium spp. Reveal the molecular basis for their recognition of “bulky-bulky” ketones. Top. Catal. 57, 356–365. doi:10.1007/s11244-013-0191-2

Mandrich, L., Merone, L., Manco, G., 2010. Hyperthermophilic phosphotriesterases/lactonases for the environment and human health. Environ. Technol. 31, 1115–27. doi:10.1080/09593331003789529

Marçal, D., Rêgo, A.T., Carrondo, M.A., Enguita, F.J., 2009. 1,3-Propanediol dehydrogenase from Klebsiella pneumoniae: Decameric quaternary structure and possible subunit cooperativity. J. Bacteriol. 191, 1143–1151. doi:10.1128/JB.01077-08

Marques, M.P.C., de Carvalho, C.C.C.R., Cabral, J.M.S., Fernandes, P., 2010. Scaling-up of complex whole-cell bioconversions in conventional and non-conventional media. Biotechnol. Bioeng. 106, 619–26. doi:10.1002/bit.22711

May, O., 2009. Green Chemistry with Biocatalysis for Production of Pharmaceuticals, in: Tao, J., Lin, G.-Q., L, A. (Eds.), Biocatalysis for the Pharmaceutical Industry: Discovery, Development and Manufacturing. John Wiley & Sons Asia (Pte) Ltd.

Medema, M.H., Breitling, R., Bovenberg, R., Takano, E., 2010. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. Nat. Rev. Microbiol. 9, 131–137. doi:10.1038/nrmicro2478

Miller, E.N., Turner, P.C., Jarboe, L.R., Ingram, L.O., 2010. Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing Escherichia coli LY180. Biotechnol. Lett. 32, 661–7. doi:10.1007/s10529-010-0209-9

Minteer, S.D. (Ed.), 2011. Enzyme Stabilization and Immobilization - Methods and Protocols, Methods in molecular biology. Humana Press, Totowa, NJ. doi:10.1007/978-1-60761-895-9

Mitchell, W., 2011. Natural products from synthetic biology. Curr. Opin. Chem. Biol. 15, 505–515. doi:10.1016/j.cbpa.2011.05.017

Moen, S.O., Fairman, J.W., Barnes, S.R., Sullivan, A., Nakazawa-Hewitt, S., Van Voorhis, W.C., Staker, B.L., Lorimer, D.D., Myler, P.J., Edwards, T.E., 2015. Structures of prostaglandin F

synthase from the protozoa Leishmania major and Trypanosoma cruzi with NADP. Acta Crystallogr. Sect. F Struct. Biol. Commun. 71, 609–614. doi:10.1107/S2053230X15006883

Mohammed Fayaz, A., Girilal, M., Rahman, M., Venkatesan, R., Kalaichelvan, P.T., 2011. Biosynthesis of silver and gold nanoparticles using thermophilic bacterium Geobacillus stearothermophilus. Process Biochem. 46, 1958–1962. doi:10.1016/j.procbio.2011.07.003

Montella, C., Bellsolell, L., Perez-Luque, R., Badia, J., Baldoma, L., Coll, M., Aguilar, J., 2005. Crystal Structure of an Iron-Dependent Group III Dehydrogenase That Interconverts L-Lactaldehyde and L-1,2-Propanediol in Escherichia coli. J. Bacteriol. 187, 4957–4966. doi:10.1128/JB.187.14.4957-4966.2005

Moon, J.-H., Lee, H.-J., Park, S.-Y., Song, J.-M., Park, M.-Y., Park, H.-M., Sun, J., Park, J.-H., Kim, B.Y., Kim, J.-S., 2011. Structures of iron-dependent alcohol dehydrogenase 2 from Zymomonas mobilis ZM4 with and without NAD+ cofactor. J. Mol. Biol. 407, 413–24. doi:10.1016/j.jmb.2011.01.045

Moore, J.C., Pollard, D.J., Kosjek, B., Devine, P., 2007. Advances in the enzymatic reduction of ketones. Acc. Chem. Res. 40, 1412–1419.

Moses, J.E., Moorhouse, A.D., 2007. The growing applications of click chemistry. Chem. Soc. Rev. 36, 1249–1262. doi:10.1039/b613014n

Murphy, A.C., 2011. Metabolic engineering is key to a sustainable chemical industry. Nat. Prod. Rep. 1406–1425. doi:10.1039/c1np00029b

Musa, M.M., Phillips, R.S., 2011. Recent advances in alcohol dehydrogenase-catalyzed asymmetric production of hydrophobic alcohols. Catal. Sci. Technol. 1, 1311–1323. doi:10.1039/c1cy00160d

Musa, M.M., Ziegelmann-Fjeld, K.I., Vieille, C., Phillips, R.S., 2008. Activity and selectivity of W110A secondary alcohol dehydrogenase from Thermoanaerobacter ethanolicus in organic solvents and ionic liquids: mono- and biphasic media. Org. Biomol. Chem. 6, 887–92. doi:10.1039/b717120j

Nandagopal, N., Elowitz, M.B., 2011. Synthetic Biology: Integrated Gene Circuits. Science 333, 1244–1248. doi:10.1126/science.1207084

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E. V, Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V. V, Osipov, G.A., Belyaev, S.S., Ivanov, M. V, 2001. Taxonomic study of aerobic thermophilic bacilli : descriptions of Geobacillus subterraneus gen . nov ., sp . nov . and Geobacillus uzenensis sp . nov . from petroleum reservoirs and transfer of Bacillus stearothermophilus , Bacillus thermocatenulatus , Ba. Int. J. Syst. Evol. Microbiol. 51, 433–446.

Neale, A.D., Scopes, R.K., Kelly, J.M., Wettenhall, R.E.H., 1986. The two alcohol dehydrogenases of Zymomonas mobilis. Eur. J. … 154, 119–124.

Neumann, H., Neumann-Staubitz, P., 2010. Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. Appl. Microbiol. Biotechnol. 87, 75–86. doi:10.1007/s00253-010-2578-3

Ni, Y., Li, C.X., Ma, H.M., Zhang, J., Xu, J.H., 2011. Biocatalytic properties of a recombinant aldo-keto reductase with broad substrate spectrum and excellent stereoselectivity. Appl. Microbiol. Biotechnol. 89, 1111–1118. doi:10.1007/s00253-010-2941-4

Nicholas, K.B., Nicholas, H.B., Deerfield, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. EMBnet News 4, 4. doi:citeulike-article-id:3198041

Niehaus, F., Bertoldo, C., Kähler, M., Antranikian, G., 1999. Extremophiles as a source of novel enzymes for industrial application. Appl. Microbiol. Biotechnol. 51, 711–29.

OECD Directorate of Science Technology and Industry, 2011. Industrial Biotechnology and Climate Change - Opportunities and Challenges [WWW Document]. URL http://www.oecd.org/sti/biotech/reportonindustrialbiotechnologyandclimatechangeopportunitiesandchallenges.htm (accessed 8.16.16).

Ogino, H., Ishikawa, H., 2001. Enzymes which are stable in the presence of organic solvents. J. Biosci. Bioeng. 91, 109–16.

Oliveira, T.B. de, Gomes, E., Rodrigues, A., 2015. Thermophilic fungi in the new age of fungal taxonomy. Extremophiles 19, 31–37. doi:10.1007/s00792-014-0707-0

Paddon, C.J., Keasling, J.D., 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Micro 12, 355–367. doi:10.1038/nrmicro3240

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., Newman, J.D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–32. doi:10.1038/nature12051

Pan, Y., Weng, J., Kabaleeswaran, V., Li, H., Cao, Y., Bhosle, R.C., Zhou, M., 2008. Cortisone dissociates the Shaker family K+ channels from their beta subunits. Nat. Chem. Biol. 4, 708–714. doi:10.1038/nchembio.114

Patel, M., Domburg, V., Hermann, B., Roes, L., Husing, B., Overbeek, L., Terragni, F., Recchia, E., 2006. Medium and Long-term Opportunities and Risks of the Biotechnological Production of Bulk Chemicals from Renewable Resources - The Potential of White Biotechnology - Final Report, European Commission GROWTH Programme [WWW Document]. URL http://dspace.library.uu.nl/bitstream/handle/1874/21824/NWS-E-2006-146.pdf?sequence=1 (accessed 8.16.16).

Pennacchio, A., Giordano, A., Rossi, M., Raia, C.A., 2011. Asymmetric Reduction of α-Keto Esters with Thermus thermophilus NADH-Dependent Carbonyl Reductase using Glucose Dehydrogenase and Alcohol Dehydrogenase for Cofactor Regeneration. European J. Org. Chem. 2011, 4361–4366. doi:10.1002/ejoc.201100107

Pennacchio, A., Rossi, M., Raia, C. A., 2013a. Synthesis of cinnamyl alcohol from cinnamaldehyde with Bacillus stearothermophilus alcohol dehydrogenase as the isolated enzyme and in recombinant E. coli cells. Appl. Biochem. Biotechnol. 170, 1482–90. doi:10.1007/s12010-013-0282-3

Pennacchio, A., Sannino, V., Sorrentino, G., Rossi, M., Raia, C.A., Esposito, L., 2013b. Biochemical and structural characterization of recombinant short-chain NAD(H)-dependent dehydrogenase/reductase from Sulfolobus acidocaldarius highly enantioselective on diaryl diketone benzil. Appl. Microbiol. Biotechnol. 97, 3949–64. doi:10.1007/s00253-012-4273-z

Peplow, M., 2016. Synthetic malaria drug meets market resistance. Nature 530, 389–390.

Peplow, M., 2013. Malaria drug made in yeast causes market ferment. Nature 494, 160–161. doi:http://dx.doi.org/10.1038/494160a

Persson, B., Hedlund, J., Jörnvall, H., 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families : the MDR superfamily. Cell. Mol. Life Sci. 65, 3879–94. doi:10.1007/s00018-008-8587-z

Philp, J.C., Ritchie, R.J., Allan, J.E.M., 2013. Biobased chemicals: the convergence of green chemistry with industrial biotechnology. Trends Biotechnol. 31, 219–222. doi:10.1016/j.tibtech.2012.12.007

Pick, A., Rühmann, B., Schmid, J., Sieber, V., 2013. Novel CAD-like enzymes from Escherichia coli K-12 as additional tools in chemical production. Appl. Microbiol. Biotechnol. 97, 5815–24. doi:10.1007/s00253-012-4474-5

Pirie, C.M., De Mey, M., Prather, K.L.J., Ajikumar, P.K., 2013. Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. ACS Chem. Biol. 8, 662–72. doi:10.1021/cb300634b

Prather, K.L.J., Martin, C.H., 2008. De novo biosynthetic pathways: rational design of microbial chemical factories. Curr. Opin. Biotechnol. 19, 468–74. doi:10.1016/j.copbio.2008.07.009

Radianingtyas, H., Wright, P.C., 2003. Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. FEMS Microbiol. Rev. 27, 593–616. doi:10.1016/S0168-6445(03)00068-8

Ran, N., Zhao, L., Chen, Z., Tao, J., 2008. Recent applications of biocatalysis in developing green chemistry for chemical synthesis at the industrial scale. Green Chem. 10, 361. doi:10.1039/b716045c

Rastogi, G., Bhalla, A., Adhikari, A., Bischoff, K.M., Hughes, S.R., Christopher, L.P., Sani, R.K., 2010. Characterization of thermostable cellulases produced by Bacillus and Geobacillus strains. Bioresour. Technol. 101, 8798–806. doi:10.1016/j.biortech.2010.06.001

Ren, Y., Strobel, G., Sears, J., Park, M., 2010. Geobacillus sp., a thermophilic soil bacterium producing volatile antibiotics. Microb. Ecol. 60, 130–6. doi:10.1007/s00248-009-9630-9

Rocha-Martín, J., Vega, D., Bolivar, J.M., Hidalgo, A., Berenguer, J., Guisán, J.M., López-Gallego, F., 2012. Characterization and further stabilization of a new anti-prelog specific alcohol dehydrogenase from Thermus thermophilus HB27 for asymmetric reduction of carbonyl compounds. Bioresour. Technol. 103, 343–50. doi:10.1016/j.biortech.2011.10.018

Rocha-Martin, J., Vega, D.E., Cabrera, Z., Bolivar, J.M., Fernandez-Lafuente, R., Berenguer, J., Guisan, J.M., 2009. Purification, immobilization and stabilization of a highly enantioselective alcohol dehydrogenase from Thermus thermophilus HB27 cloned in E. coli. Process Biochem. 44, 1004–1012. doi:10.1016/j.procbio.2009.04.026

Roemer, T., Davies, J., Giaever, G., Nislow, C., 2011. Bugs, drugs and chemical genomics. Nat. Chem. Biol. 8, 46–56. doi:10.1038/nchembio.744

Ruder, W.C., Lu, T., Collins, J.J., 2011. Synthetic Biology Moving into the Clinic. Science 333, 1248–1252. doi:10.1126/science.1206843

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945. doi:10.1093/bioinformatics/16.10.944

Salamanova, E., Tsoneva, D., Karshikoff, A., 2013. Physical bases of thermal stability of proteins: A comparative study on homologous pairs from mesophilic and thermophilic organisms 45, 592–600.

Sammond, D.W., Kastelowitz, N., Himmel, M.E., Yin, H., Crowley, M.F., Bomble, Y.J., 2016. Comparing residue clusters from thermophilic and mesophilic enzymes reveals adaptive mechanisms. PLoS One 11, 1–18. doi:10.1371/journal.pone.0145848

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., Hughes, G.J., 2010. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 329, 305–9. doi:10.1126/science.1188934

Schroer, K., Tacha, E., Lütz, S., 2007. Process Intensification for Substrate-Coupled Whole Cell Ketone Reduction by In Situ Acetone Removal. Org. Process Res. Dev. 11, 836–841.

Serrano-Ruiz, J.C., Luque, R., Sepúlveda-Escribano, A., 2011. Transformations of biomass-derived platform molecules: from high added-value chemicals to fuels via aqueous-phase processing. Chem. Soc. Rev. 40, 5266–81. doi:10.1039/c1cs15131b

Sharp, P.A., Cooney, C.L., Kastner, M.A., Lees, J., Sasisekharan, R., Yaffe, M.B., Bhatia, S.N., Jacks, T.E., Lauffenburger, D.A., Langer, R., Hammond, P.T., Sur, M., 2011. The Third Revolution : The Convergence of the Life Sciences , Physical Sciences , and Engineering [WWW Document]. URL https://www.cimit.org/images/about/MIT-White-Paper-on-Convergence.pdf

Sheldon, R.A., 2012. Fundamentals of green chemistry: efficiency in reaction design. Chem. Soc. Rev. 41, 1437–51. doi:10.1039/c1cs15219j

Shih, T.-W., Pan, T.-M., 2011. Stress responses of thermophilic Geobacillus sp. NTU 03 caused by heat and heat-induced stress. Microbiol. Res. 166, 346–59. doi:10.1016/j.micres.2010.08.001

Shin, J.H., Kim, H.U., Kim, D.I., Lee, S.Y., 2013. Production of bulk chemicals via novel metabolic pathways in microorganisms. Biotechnol. Adv. 31, 925–35. doi:10.1016/j.biotechadv.2012.12.008

Siddiqui, K.S., 2015. Some like it hot , some like it cold : Temperature dependent biotechnological applications and improvements in extremophilic enzymes. Biotechnol. Adv. 33, 1912–1922. doi:10.1016/j.biotechadv.2015.11.001

Simpson, H., Cowan, D., 1992. Secondary alcohol dehydrogenases from extremely thermophilic bacteria. Ann. N. Y. Acad. Sci. 145–151.

Singh, R., Soni, P., Kumar, P., Purohit, S., Singh, A., 2008. Biodegradation of high explosive production effluent containing RDX and HMX by denitrifying bacteria. World J. Microbiol. Biotechnol. 25, 269–275. doi:10.1007/s11274-008-9889-x

Smith, M., Chen, B., Hibbert, E., Kaulmann, U., Smithies, K., Galman, J., Baganz, F., Dalby, P., Hailes, H., Lye, G., Ward, J., Woodley, J., Micheletti, M., 2010. A Multidisciplinary Approach Toward the Rapid and Preparative-Scale Biocatalytic Synthesis of Chiral Amino Alcohols: A Concise Transketolase-/omega-Transaminase-Mediated Synthesis of (2S,3S)-2-Aminopentane-1,3-diol 14, 99–107.

Spickermann, D., Kara, S., Barackov, I., Hollmann, F., Schwaneberg, U., Duenkelmann, P., Leggewie, C., 2014. Alcohol dehydrogenase stabilization by additives under industrially relevant reaction conditions. J. Mol. Catal. B Enzym. 103, 24–28.

doi:10.1016/j.molcatb.2013.11.015

Stahel, W.R., 2016. Circular Economy. Nature 531, 435–438.

Stephanopoulos, G., 2012. Synthetic biology and metabolic engineering. ACS Synth. Biol. 1, 514–25. doi:10.1021/sb300094q

Stephenson, R., Stuart, J., 1986. Mutual binary solubilities: water-alcohols and water-esters. J. Chem. Eng. Data 56–70.

Stephenson, R., Stuart, J., Tabak, M., 1984. Mutual solubility of water and aliphatic alcohols. J. Chem. Eng. Data 287–290.

Stephenson, R.M., 1993. Mutual solubility of water and aldehydes. J. Chem. Eng. Data 38, 630–633. doi:10.1021/je00012a040

Sterpone, F., Melchionna, S., 2012. Thermophilic proteins: insight and perspective from in silico experiments. Chem. Soc. Rev. 41, 1665–76. doi:10.1039/c1cs15199a

Straathof, A.J.J., Panke, S., Schmid, A., 2002. The production of fine chemicals by biotransformations. Curr. Opin. Biotechnol. 13, 548–56.

Sulzenbacher, G., Alvarez, K., Van Den Heuvel, R.H.H., Versluis, C., Spinelli, S., Campanacci, V., Valencia, C., Cambillau, C., Eklund, H., Tegoni, M., 2004. Crystal structure of E. coli alcohol dehydrogenase YqhD: Evidence of a covalently modified NADP coenzyme. J. Mol. Biol. 342, 489–502. doi:10.1016/j.jmb.2004.07.034

Takami, H., Nishi, S., Lu, J., Shimamura, S., Takaki, Y., 2004a. Genomic characterization of thermophilic Geobacillus species isolated from the deepest sea mud of the Mariana Trench. Extremophiles 8, 351–6. doi:10.1007/s00792-004-0394-3

Takami, H., Takaki, Y., Chee, G.-J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S., Uchiyama, I., 2004b. Thermoadaptation trait revealed by the genome sequence of thermophilic Geobacillus kaustophilus. Nucleic Acids Res. 32, 6292–303. doi:10.1093/nar/gkh970

Tamarit, J., Cabiscol, E., Aguilar, J., Ros, J., 1997. Differential inactivation of alcohol dehydrogenase isoenzymes in Zymomonas mobilis by oxygen. J. Bacteriol. 179.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–9. doi:10.1093/molbev/mst197

Tang, Y.J., Sapra, R., Joyner, D., Hazen, T.C., Myers, S., Reichmuth, D., Blanch, H., Keasling, J.D., 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant Geobacillus strain. Biotechnol. Bioeng. 102, 1377–86. doi:10.1002/bit.22181

Taylor, M.P., Esteban, C.D., Leak, D.J., 2008. Development of a versatile shuttle vector for gene expression in Geobacillus spp. Plasmid 60, 45–52. doi:10.1016/j.plasmid.2008.04.001

Tietze, L.F., 1996. Domino Reactions in Organic Synthesis. Chem. Rev. 96, 115–136.

Tietze, L.F., Modi, A., 2000. Multicomponent domino reactions for the synthesis of biologically active natural products and drugs. Med. Res. Rev. 20, 304–22.

Timasheff, S.N., 1992. Solvent effects on protein stability. Curr. Opin. Struct. Biol. 2, 35–39.

Tipton, K.F., Armstrong, R.N., Bakker, B.M., Bairoch, A., Cornish-Bowden, A., Halling, P.J.,

Hofmeyr, J.-H., Leyh, T.S., Kettner, C., Raushel, F.M., Rohwer, J., Schomburg, D., Steinbeck, C., 2014. Standards for Reporting Enzyme Data: The STRENDA Consortium: What it aims to do and why it should be helpful. Perspect. Sci. 1, 131–137. doi:10.1016/j.pisc.2014.02.012

Totir, M., Echols, N., Nanao, M., Gee, C.L., Moskaleva, A., Gradia, S., Iavarone, A.T., Berger, J.M., May, A.P., Zubieta, C., Alber, T., 2012. Macro-to-micro structural proteomics: Native source proteins for high-throughput crystallization. PLoS One 7, 1–12. doi:10.1371/journal.pone.0032498

Tourova, T.P., Korshunova, A. V., Mikhailova, E.M., Sokolova, D.S., Poltaraus, A.B., Nazina, T.N., 2010. Application of gyrB and parE sequence similarity analyses for differentiation of species within the genus Geobacillus. Microbiology 79, 356–369. doi:10.1134/S0026261710030124

Tourova, T.P., Nazina, T.N., Mikhailova, E.M., Rodionova, T.A., Ekimov, A.N., Mashukova, A. V., Poltaraus, A.B., 2008. alkB homologs in thermophilic bacteria of the genus Geobacillus. Mol. Biol. 42, 217–226. doi:10.1134/S0026893308020076

Tripathy, S., Maiti, N.K., 2014. Construction of Geobacillus thermoglucosidasius cDNA library and analysis of genes expressed in response to heat stress. Mol. Biol. Rep. 41, 1639–44. doi:10.1007/s11033-013-3011-7

Tse, P., Scopes, R.K., Wedd, A.G., 1989. Iron-activated alcohol dehydrogenase from Zymomonas mobilis: isolation of apoenzyme and metal dissociation constants. J. Am. Chem. Soc. 111, 8703–8706. doi:10.1021/ja00205a021

Tsuruta, H., Paddon, C.J., Eng, D., Lenihan, J.R., Horning, T., Anthony, L.C., Regentin, R., Keasling, J.D., Renninger, N.S., Newman, J.D., 2009. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. PLoS One 4, e4489. doi:10.1371/journal.pone.0004489

University College London, 2014. Impact Case Study (REF3b), Research Excellence Framework 2014 [WWW Document]. URL http://impact.ref.ac.uk/casestudies2/refservice.svc/GetCaseStudyPDF/36403 (accessed 8.16.16).

Vennestrøm, P.N.R., Osmundsen, C.M., Christensen, C.H., Taarning, E., 2011. Beyond petrochemicals: the renewable chemicals industry. Angew. Chem. Int. Ed. Engl. 50, 10502–9. doi:10.1002/anie.201102117

Vieille, C., Zeikus, G.J., Vieille, C., 2001. Hyperthermophilic Enzymes : Sources , Uses , and Molecular Mechanisms for Thermostability Hyperthermophilic Enzymes : Sources , Uses , and Molecular Mechanisms for Thermostability. Microbiol. Mol. Biol. Rev. 65, 1–43. doi:10.1128/MMBR.65.1.1

Vincze, T., Posfai, J., Roberts, R.J., 2003. NEBcutter: A program to cleave DNA with restriction enzymes. Nucleic Acids Res. 31, 3688–3691. doi:10.1093/nar/gkg526

Welsh, F., 1989. Microbiological and enzymatic production of flavor and fragrance chemicals. Crit. Rev. Biotechnol. 9, 105–169.

Wenda, S., Illner, S., Mell, A., Kragl, U., 2011. Industrial biotechnology—the future of green chemistry? Green Chem. 3007–3047. doi:10.1039/c1gc15579b

Weng, J., Cao, Y., Moss, N., Zhou, M., 2006. Modulation of voltage-dependent shaker family potassium channels by an aldo-keto reductase. J. Biol. Chem. 281, 15194–15200.

doi:10.1074/jbc.M513809200

Werpy, T.A., Holladay, J.E., White, J.F., 2004. Top Value Added Chemicals From Biomass: I. Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Richland, WA. doi:10.2172/926125

Wettstein, S.G., Alonso, D.M., Gürbüz, E.I., Dumesic, J.A., 2012. A roadmap for conversion of lignocellulosic biomass to chemicals and fuels. Curr. Opin. Chem. Eng. doi:10.1016/j.coche.2012.04.002

Willies, S., Isupov, M., Littlechild, J., 2010. Thermophilic enzymes and their applications in biocatalysis: a robust aldo-keto reductase. Environ. Technol. 31, 1159–67. doi:10.1080/09593330.2010.490857

Wohlgemuth, R., 2009. The locks and keys to industrial biotechnology. N. Biotechnol. 25, 204–13. doi:10.1016/j.nbt.2009.01.002

Wu, J.T., Wu, L.H., Knight, J.A., 1986. Stability of NADPH: effect of various factors on the kinetics of degradation. Clin. Chem. 32, 314–9.

Wu, X., Kobori, H., Orita, I., Zhang, C., Imanaka, T., Xing, X.-H., Fukui, T., 2012. Application of a novel thermostable NAD(P)H oxidase from hyperthermophilic archaeon for the regeneration of both NAD^+ and $NADP^+$. Biotechnol. Bioeng. 109, 53–62. doi:10.1002/bit.23294

Xu, Y.-P., Guan, Y.H., Yu, H.-L., Ni, Y., Ma, B.-D., Xu, J.-H., 2014. Improved o-chlorobenzoylformate bioreduction by stabilizing aldo-keto reductase YtbE with additives. J. Mol. Catal. B Enzym. 104, 108–114. doi:10.1016/j.molcatb.2014.03.006

Yan, Y., Lee, C.-C., Liao, J.C., 2009. Enantioselective synthesis of pure (R,R)-2,3-butanediol in Escherichia coli with stereospecific secondary alcohol dehydrogenases. Org. Biomol. Chem. 7, 3914–7. doi:10.1039/b913501d

Yang, H., Jonsson, A., Wehtje, E., Adlercreutz, P., Mattiasson, B., 1997. The enantiomeric purity of alcohols formed by enzymatic reduction of ketones can be improved by optimisation of the temperature and by using a high co-substrate concentration. Biochim. Biophys. Acta - Gen. Subj. 1336, 51–58. doi:10.1016/S0304-4165(97)00010-X

Yang, J.Y., Karr, J.R., Watrous, J.D., Dorrestein, P.C., 2011. Integrating "-omics" and natural product discovery platforms to investigate metabolic exchange in microbiomes. Curr. Opin. Chem. Biol. 15, 79–87. doi:10.1016/j.cbpa.2010.10.025

Yildirim, M., Colak, A., Col, M., Canakci, S., 2009. A new recombinant phosphotriesterase homology protein from Geobacillus caldoxylosilyticus TK4: An extremely thermo- and pH-stable esterase. Process Biochem. 44, 1366–1373. doi:10.1016/j.procbio.2009.07.014

Ying, X., Grunden, A.M., Nie, L., Adams, M.W.W., Ma, K., 2009. Molecular characterization of the recombinant iron-containing alcohol dehydrogenase from the hyperthermophilic Archaeon, Thermococcus strain ES1. Extremophiles 13, 299–311. doi:10.1007/s00792-008-0217-z

Ying, X., Ma, K., 2011. Characterization of a zinc-containing alcohol dehydrogenase with stereoselectivity from the hyperthermophilic archaeon Thermococcus guaymasensis. J. Bacteriol. 193, 3009–3019. doi:10.1128/JB.01433-10

Zambare, V.P., Bhalla, A., Muthukumarappan, K., Sani, R.K., Christopher, L.P., 2011. Bioprocessing of agricultural residues to ethanol utilizing a cellulolytic extremophile.

Extremophiles 15, 611–8. doi:10.1007/s00792-011-0391-2

Zeigler, D.R., 2014. The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? Microbiology 160, 1–11. doi:10.1099/mic.0.071696-0

Zeigler, D.R., 2005. Application of a recN sequence similarity analysis to the identification of species within the bacterial genus Geobacillus. Int. J. Syst. Evol. Microbiol. 55, 1171–9. doi:10.1099/ijs.0.63452-0

Zeldes, B.M., Keller, M.W., Loder, A.J., Straub, C.T., Adams, M.W.W., Kelly, R.M., 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front. Microbiol. 6, 1–17. doi:10.3389/fmicb.2015.01209

Zhang, F., Rodriguez, S., Keasling, J.D., 2011. Metabolic engineering of microbial pathways for advanced biofuels production. Curr. Opin. Biotechnol. 22, 775–83. doi:10.1016/j.copbio.2011.04.024

Zhang, X.-Z., Zhang, Y.-H.P., 2010. One-step production of biocommodities from lignocellulosic biomass by recombinant cellulolytic Bacillus subtilis: Opportunities and challenges. Eng. Life Sci. 10, 398–406. doi:10.1002/elsc.201000011

Zhang, Y.-H.P., Sun, J., Zhong, J.-J., 2010. Biofuel production by in vitro synthetic enzymatic pathway biotransformation. Curr. Opin. Biotechnol. 21, 663–9. doi:10.1016/j.copbio.2010.05.005

Zhou, C.-H.C., Beltramini, J.N., Fan, Y.-X., Lu, G.Q.M., 2008. Chemoselective catalytic conversion of glycerol as a biorenewable source to valuable commodity chemicals. Chem. Soc. Rev. 37, 527–49. doi:10.1039/b707343g

Ziegelmann-Fjeld, K.I., Musa, M.M., Phillips, R.S., Zeikus, J.G., Vieille, C., 2007. A Thermoanaerobacter ethanolicus secondary alcohol dehydrogenase mutant derivative highly active and stereoselective on phenylacetone and benzylacetone. Protein Eng. Des. Sel. 20, 47–55. doi:10.1093/protein/gzl052

7. APPENDIX ONE – REFERENCE DATA

This section contains a record of all additional information with regards to the genes, enzymes, primers and designations used in this thesis.

7.1. ALTERNATIVE DESIGNATIONS OF GENES AND ENZYMES

Below is a list of all external designations utilised during this project. For simplicity all enzymes were designated simply by a letter during this thesis; however, there are multiple differing designations given by external sources assigned to each gene and gene product, depending on the database. Further, internal designations were used during the course of the project to ensure consistency, which are also below. NCBI refers to the National Center for Biotechnology Information databases, referencing the public *G. thermoglucosidasius* genome designated C56-YS93, RAST refers to the Rapid Annotation Using Subsystem Technology database (Aziz et al., 2008), referencing the private genome TM242, and RTMO numbers refer to a previous version of the same genome. Wherever relevant, TM242 was considered the primary source for sequencing purposes.

ADH A Alternative Designations (Full Gene/Enzyme):

- RTMO 02838
- RAST 3856
- NCBI Reference (Gene): Geoth_3897
- NCBI Reference (Protein): YP_004589790.1

ADH B Alternative Designations:

- RTMO 04745
- RAST 240
- NCBI Reference (Gene): Geoth_3237
- NCBI Reference (Protein): YP_004589190.1

ADH C Alternative Designations:

- RTMO 01825
- RAST 3939
- NCBI Reference (Gene): Geoth_3827
- NCBI Reference (Protein): YP_004589720.1

ADH D Alternative Designations:

- RTMO 01643
- RAST 1340
- NCBI Reference (Gene): Geoth_2331
- NCBI Reference (Protein): YP_004588337.1
- Internal Designation: ADH 2

ADH E Alternative Designations

- RTMO 03537
- RAST 1773
- NCBI Reference (Gene): Geoth_1941
- NCBI Reference (Protein): YP_004587981.1
- Internal Designation: ADH 6

ADH F Alternative Designations

- RTMO 04757
- RAST 208
- NCBI Reference (Gene): Geoth_3301
- NCBI Reference (Protein): YP_004589250.1
- Internal Designation: ADH 10

ADH G Alternative Designations

- RTMO 03137
- RAST 3154
- NCBI Reference (Gene): Geoth_0631
- NCBI Reference (Protein): YP_004586722.1
- Internal Designation: ADH 11

7.2. LIST OF ALL PRIMERS

Below is a list of all primers used to isolate genes encoding the enzymes referenced in this work.

Gene	Sequence
ADH1 F	GGGGGGGCTAGCAGTGTTTCCCGTATTGTATTTACATCGCTG
ADH1 R	GGGGGGCTCGAGTTATTTCGGCTCCGTTACCACACC
ADH2 F	GGGGGGGCTAGCAAATATCGCAAACTTGGAAGAACGG
ADH2 R	GGGGGGCTCGAGTTATGCAAGAATTTGTTCAATTTTCGCAAG
ADH3 F	GGGGGGGCTAGCAAAACTGCGAAAATTGTTCAGCACAAA
ADH3 R	GGGGGGCTCGAGTTAGAATTTAGTGATTACGTTAAACCCTTGGT
ADH4 F	GGGGGGGCTAGCAAAGCGGCAGTTGTAACGAGTTCAAA
ADH4 R	GGGGGGCTCGAGTTAACGATTGACACCGATGGTTAAAACG
ADH5 F	GGGGGGGCTAGCAAAGCGTTGGTCAAAAAAGAATTGGGTTTC
ADH5 R	GGGGGGCTCGAGTTAGTCGATAGGAGTTAACAATACTTTGATTGCCTCTC
ADH6 F	ACACACGCTAGCAATACATTCTTCTTGAAACCAAAAATCTACTTCGGAAA
ADH6 R	ACACACCTCGAGTTACCAATCTTATCCGTTATATGCCCATTT
ADH7 F	CTCTCTGCTAGCGATGTTATTGCTATTACTGGGGCG
ADH7 R	CTCTCTCTCGAGTTAAGATAAAATCATTTCGTTTTGACCG
ADH8 F	GAAAGAGCTAGCGTCACGAAACAGCAGGCAAC
ADH8 R	AAGAGGCTCGAGTTACCCATTCACCACTTTCCCAC
ADH9 F	GGGGGGGCTAGCAAGTTTTCACAAAAAGTGGTTGTCGTAA
ADH9 R	GGGGGGCTCGAGTTATTCGGCGTAAATCATCTTTCTTGTC
ADH10 F	GGGGGGGCTAGCAAACATCTGCAAGACTGTGTCACACTACAT
ADH10 R	GGGGGGCTCGAGTTAAAAATCAAAATTGTCTGGGTCCG
ADH11 F	GGGGTTGCTAGCCTTACTGAAAATATACATAAGTTTGTGAAA
ADH11 R	GGGGTTCTCGAGTTACAACGATGCACGCA
ADH12 F	GGGGGGGCTAGCATCCGATTTGCCACCATC
ADH12 R	GGGGGGCTCGAGTTATTGGTCAGCAGGAAATACGAT
ADH13 F	CCCCCCGCTAGCCGCAAAAAGAAGCAGATTACCCG
ADH13 R	CCCCCCCTCGAGTTATGGAACCTCATGACCGACTGTG
ADH15 F	GGGGGGGCTAGCAAAAAACGGCGCATCGGCTTGT
ADH15 R	GGGGGGCTCGAGTTAGCGGTGCATGTTGTACACTGAAG

Table 35 – List of all successful primers used to isolate genes encoding putative ADHs. Forward primers use *Nhe*I and reverse primers use *Xho*I as restriction enzymes. Primers have six bases before the restriction enzyme site, typically poly G, but this section may be altered to remove hairpins or dimers. Underlined bases indicate restriction enzyme sites.

7.3. FULL SEQUENCES OF GENES AND ENZYMES

The full gene and protein sequences of all enzymes referred to in Chapters 2 and 3 are given below.

7.3.1. ADH A (FULL GENE/ENZYME)

atggctgtggaggagagagtcgtcgataaaaaaatcgaagtagcaaaaatgattgatgagcttg
tcgctaatgcacagaaagcgttggaacaaattcgcgcttacgatcaagaaacgatcgatcatat
cgtgaaagaaatggcgttagccgggctcgacaagcatatggcattagccaagcttgcagtagaa
gaaacaaaacgcggtgtatatgaagataaaatcataaaaaaccttttgcgacagaatatatat
accacaatattaagtatgataaaacagtcgggattattcatgaaaatccgcatgaagaaattat
cgaaattgctgagcctgttggtgttattgctgggattacgccagtgacaaacccgacatcgaca
acgatgtttaaagcgttaatctcgataaaaacacgcaacccgattattttcgctttccatccat
cggcgcaacgatgcagcagcgaagcggcaagagtgctgcgcgatgcggcggtccgggcaggggc
tccagaacattgcattcaatggattgaaactccttcgcttgatgcaaccaatcagcttatgcac
catcctggcgtttctctcattttggcaactggtggcgccggcatggtgaaagcagcgtacagct
ctggaaaaccagctttgggcgtcggacctggcaatgtgccttgctatattgaaaaaacggcaaa
cataaaacgggcggtaaatgacttaattttatcgaaaacgtttgataacggcatgatttgcgct
tctgaacaagcagtcattattgataaagaaatttatgaacaagtaaagaaagaaatgatagaaa
accattgttatttcttaaatgaagaagaaaagaaaaaagtagaaaaactcgttatcaatgaaaa
tacatgcgccgtcaacccggatatcgtcggaaagccagcttatgaaattgcgaaaatggccggc
atcgctgtgccggaagacacaaaaattcttgttgctgagttaaaaggggtcgggccaaaatatc
cgttgtctcgggaaaaattaagccctgtccttgcttgctataaagttaacagcacggaagaagg
atttaagcgctgtgaagaaatgctggaatttggcggcttgggacattcggctgtcatccattcc
gataatcaaaacgtggttaccgaatttggcaaacggatgaaagcgggacggattatcgttaatg
cgccatcttcgcaaggagcaatcggcgatatttacaatgcgtacattccgtcattaacgctggg
atgcggcacatttggcggaaactctgtttcgacaaacgtcagtgcgattcatcttatcaatata
aaaagaatggcaaaaaggacggtaaatatgcaatggtttaaagtgccgccgaaaatttatttcg
aaaaaaatgctgtacaatacttagcgaaaatgccggatatttccagagcttttatcgtcaccga
cccgggaatggtcaagctcggatatgtcgataaagtgctgtattacttgcgcagacgcccggat
tatgtgcatagtgaaattttctccgaagtagagccagatccttcaattgagacggtaatgaaag
gtgtcgatatgatgagaagtttcgagccggatgtgattatcgcgcttggaggcggctcgccaat
ggatgcggcaaaagcgatgtggctcttttacgagcatccgacagcggatttcaacgcattaaaa
caaaaatttttagatattcgaaaacgcgtttataaatatccaaaactgggccaaaaagcgaaat
ttgtcgccattccgacgacatcaggaacaggatcggaagtaacgtcctttgccgtcattaccga
taaaaaaacgaatataaaatatccgttggcagattatgaattgacaccggacgtcgcgattgtg
gatccgcaatttgtcatgaccgtgccaaaacatgtcaccgccgatacgggaatggatgtattga
cacatgcgatcgaagcgtatgtctccaatatggcaaatgattataccgatggtcttgccatgaa
agcaatccaactcgtatttgaatatttgccgcgggcatatcaaaacggagcggatgagcttgcc
cgggagaaaatgcataacgcctctacgattgcgggaatggcatttgccaacgcgtttttaggca
ttaaccatagtttggctcataaacttggcgcggaattccatattccgcatgggcgcgcgaatac
cattttgatgccgcatgtcattcgctataacgcagcgaaaccgaaaaaatttaccgcatttccg
aaatacgaatatttcaaagcggaccagcgctatgcagaaattgcgagaatgctcggcttgccgg
cccgcacaacggaagaaggggtcgaaagcctcgttcaggcgatcattaagctggcaaaacagtt
ggatatgccgctgagcattgaagcatgcggcgtcagcaaacaagaatttgaaagcaaagttgaa
aaattagccgaattggctttcgaagaccaatgtactactgctaacccgaaactcccgcttgtta
gcgatttagttcatatttatcgccaagcgtttaaaggagtttaa

MAVEERVVDKKIEVAKMIDELVANAQKALEQIRAYDQETIDHIVKEMALAGLDKHMALAKLAVE
ETKRGVYEDKIIKNLFATEYIYHNIKYDKTVGIIHENPHEEIIEIAEPVGVIAGITPVTNPTST
TMFKALISIKTRNPIIFAFHPSAQRCSSEAARVLRDAAVRAGAPEHCIQWIETPSLDATNQLMH
HPGVSLILATGGAGMVKAAYSSGKPALGVGPGNVPCYIEKTANIKRAVNDLILSKTFDNGMICA
SEQAVIIDKEIYEQVKKEMIENHCYFLNEEEKKKVEKLVINENTCAVNPDIVGKPAYEIAKMAG
IAVPEDTKILVAELKGVGPKYPLSREKLSPVLACYKVNSTEEGFKRCEEMLEFGGLGHSAVIHS
DNQNVVTEFGKRMKAGRIIVNAPSSQGAIGDIYNAYIPSLTLGCGTFGGNSVSTNVSAIHLINI

KRMAKRTVNMQWFKVPPKIYFEKNAVQYLAKMPDISRAFIVTDPGMVKLGYVDKVLYYLRRRPD
YVHSEIFSEVEPDPSIETVMKGVDMMRSFEPDVIIALGGGSPMDAAKAMWLFYEHPTADFNALK
QKFLDIRKRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFAVITDKKTNIKYPLADYELTPDVAIV
DPQFVMTVPKHVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAYQNGADELA
REKMHNASTIAGMAFANAFLGINHSLAHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFP
KYEYFKADQRYAEIARMLGLPARTTEEGVESLVQAIIKLAKQLDMPLSIEACGVSKQEFESKVE
KLAELAFEDQCTTANPKLPLVSDLVHIYRQAFKGV

7.3.2. ADH A (ADH SECTION ONLY)

atgaatatgcaatggtttaaagtgccgccgaaaatttatttcgaaaaaaatgctgtacaatact
tagcgaaaatgccggatatttccagagcttttatcgtcaccgacccgggaatggtcaagctcgg
atatgtcgataaagtgctgtattacttgcgcagacgcccggattatgtgcatagtgaaatttttc
tccgaagtagagccagatccttcaattgagacggtaatgaaaggtgtcgatatgatgagaagtt
tcgagccggatgtgattatcgcgcttggaggcggctcgccaatggatgcggcaaaagcgatgtg
gctcttttacgagcatccgacagcggatttcaacgcattaaaacaaaaatttttagatattcga
aaacgcgtttataaatatccaaaactgggccaaaaagcgaaatttgtcgccattccgacgacat
caggaacaggatcggaagtaacgtcctttgccgtcattaccgataaaaaaacgaatataaaata
tccgttggcagattatgaattgacaccggacgtcgcgattgtggatccgcaatttgtcatgacc
gtgccaaaacatgtcaccgccgatacgggaatggatgtattgacacatgcgatcgaagcgtatg
tctccaatatggcaaatgattataccgatggtcttgccatgaaagcaatccaactcgtatttga
atatttgccgcgggcatatcaaaacggagcggatgagcttgcccgggagaaaatgcataacgcc
tctacgattgcgggaatggcatttgccaacgcgtttttaggcattaaccatagtttggctcata
aacttggcgcggaattccatattccgcatgggcgcgcgaataccattttgatgccgcatgtcat
tcgctataacgcagcgaaaccgaaaaaatttaccgcatttccgaaatacgaatatttcaaagcg
gaccagcgctatgcagaaattgcgagaatgctcggcttgccggcccgcacaacggaagaagggg
tcgaaagcctcgttcaggcgatcattaagctggcaaaacagttggatatgccgctgagcattga
agcatgcggcgtcagcaaacaagaatttgaaagcaaagttgaaaaattagccgaattggctttc
gaagaccaatgtactactgctaacccgaaactcccgcttgttagcgatttagttcatatttatc
gccaagcgtttaaaggagtttaa

MNMQWFKVPPKIYFEKNAVQYLAKMPDISRAFIVTDPGMVKLGYVDKVLYYLRRRPDYVHSEIF
SEVEPDPSIETVMKGVDMMRSFEPDVIIALGGGSPMDAAKAMWLFYEHPTADFNALKQKFLDIR
KRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFAVITDKKTNIKYPLADYELTPDVAIVDPQFVMT
VPKHVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAYQNGADELAREKMHNA
STIAGMAFANAFLGINHSLAHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFPKYEYFKA
DQRYAEIARMLGLPARTTEEGVESLVQAIIKLAKQLDMPLSIEACGVSKQEFESKVEKLAELAF
EDQCTTANPKLPLVSDLVHIYRQAFKGV

7.3.3. ADH B

atgaaagcacttacatacctagggccaggaaaaaaagaattaatggaaaaaccaaagccaaaaa
ttgaaaaggaaaccgatgcaatcgtcaaaataataaaaacaacgatttgtggaacggatttgca
cattctttcaggagatgttcctactgttgaagaagggcggattttaggacacgaaggcgtcgga
attattgaagaagttggttcggccgtaaagaattttaaaaaaggcgacagagtgttgatttctt
gcattacctcttgtggaaaatgcgaaaattgcaagaaagggttatacgcccattgcgaagatgg
cggctggatcttgggccacttaattgatggaactcaagcagaatatgtaagaattccgcacgca
gacaacagcctttatcctattccggaaggcgtggatgaagagactcttgtcatgcttagtgaca
ttcttccaacaggatttgaaatcggcgtgttgaacggcaaagttcagcctggacaaaccgtcgc
cattatcggagctggtcccgtaggtatggcagcgctattaacagcccaattttattcaccagca
gagatcattatggttgatttagacgataaccgtttagaagttgcgaaaaaatttggcgcgaccc
aagtggtgaatagcgctgatggcaaggcagtggaaaaaattatggaattaaccggcgggaaagg
tgtagacgtcgcgatggaagccgtcggaattccggcaacatttgatatttgtcaagaaattgtc

aaaccaggtggctatatcgccaatatcggtgttcatggaaaaagcgtggaatttcacattgaaa
aattatggatacgcaacattacgttgacaaccggtcttgtcaacacgacttctacgccgatgtt
attaaaaacggtgcagtcgaaaaaattgaagccggaacaattaattacccatcgtttcgccttt
tcagacattatgaaagcgtatgaagtatttggaaatgcagcaaaagaaaaagcgttaaaagtca
ttatttccaacagttaa

MKALTYLGPGKKELMEKPKPKIEKETDAIVKIIKTTICGTDLHILSGDVPTVEEGRILGHEGVG
IIEEVGSAVKNFKKGDRVLISCITSCGKCENCKKGLYAHCEDGGWILGHLIDGTQAEYVRIPHA
DNSLYPIPEGVDEETLVMLSDILPTGFEIGVLNGKVQPGQTVAIIGAGPVGMAALLTAQFYSPA
EIIMVDLDDNRLEVAKKFGATQVVNSADGKAVEKIMELTGGKGVDVAMEAVGIPATFDICQEIV
KPGGYIANIGVHGKSVEFHIEKLWIRNITLTTGLVNTTSTPMLLKTVQSKKLKPEQLITHRFAF
SDIMKAYEVFGNAAKEKALKVIISNS

7.3.4. ADH C

atgaaagcagcgcgatggtacaatgctagagacattcgagtggaagaagtagaagaaccgaaag
caggaaaaggaaaagtaaaaattaaagtcgaatgggcgggaatttgcggaagcgatttgcacga
atatgcggcaggtccaatttttattcctgtccaaaatcctcatcccgttagtaaagatgtcgcc
ccgattatcatgggccacgaattttcgggccgagtagtagaagtcggggaaggagttactaaag
tcaaagttggggatcctgtcgttgttgaaccgattcttcgttgtggagaatgcccggcttgcaa
aaaagggaaatacaatctttgcgaacatttaggattccatggcctgtccggaggaggcggcgga
ttctcagaatataccgtcgttgatgaatatatggtgcacaaaatgcctgaaggtctttcttttg
aacaaggagcgctagtggaaccggcggctgtcgcgttacatgctgtcagatcaagcaaaatcaa
acctggtgataaagctgccgtttttggaacaggaccgataggccttcttgtgatggaagcattg
aaagcggctggcgcctcggaaatttatgcagtcgaagtttccaaagaacgtttgcaaaaagcga
aagagcttggcgccacatctgtcatcaatccaaaagaagaagatccggttcgcaaacttgtcga
attgaccggtggcggcgtcgatgttgcgtttgaagtaacgggggtgccggccgttttgcagcaa
gctattgacagcactgcattcgaaggagaaacggttatcgttagtatatgggaaaaagaagcaa
gcattcagccaaacaatattgtattaaaagaaagaaatgtgaagggagttattgcataccgcga
tattttccctgctgtaatggaattaatgaaacgaggctatttccaagccgaaaagcttgttacg
aaacgaatcaagctggatgatatcgtggcagaaggatttgaagcgctcattaaagaaaaagacc
aagtgaaaattttggtcaaaccagaataa

MKAARWYNARDIRVEEVEEPKAGKGKVKIKVEWAGICGSDLHEYAAGPIFIPVQNPHPVSKDVA
PIIMGHEFSGRVVEVGEGVTKVKVGDPVVVEPILRCGECPACKKGKYNLCEHLGFHGLSGGGGG
FSEYTVVDEYMVHKMPEGLSFEQGALVEPAAVALHAVRSSKIKPGDKAAVFGTGPIGLLVMEAL
KAAGASEIYAVEVSKERLQKAKELGATSVINPKEEDPVRKLVELTGGGVDVAFEVTGVPAVLQQ
AIDSTAFEGETVIVSIWEKEASIQPNNIVLKERNVKGVIAYRDIFPAVMELMKRGYFQAEKLVT
KRIKLDDIVAEGFEALIKEKDQVKILVKPE

7.3.5. ADH D

atgaaataccggaagcttggaagaacggggttaaaagtaagtgaaattagtttaggcagctggc
ttacatatggcaattcggttgaaaaagaaacggcgatccgcgtcattgataaagcgtatgagtt
gggaatcaactcgtttgacacggcaaatgtgtacgcgaaaggggaagcggaaaaaatcgtcggc
gaagcgttgcgcaagtatccgcgcgaatcatacgtattggcgacgaaagtgtattggccgatgg
gggacggcccaaacgaccgcggcttgtcgcggaaacatgtgtttgagcagcttcacgccagctt
gaagcgcctgcagttagattacgttgatatttattattgtcatcgctatgatagtgaaacaccg
atcgatgaaacattgcgcacgattgatgaccttgtccgtcaaggaaaagtattgtatgtcggtg
tcagtgaatggacggcccagcaaattcaagaagcgcttggcacggccgataaatatttattgga
ccgcatcgtcgtcaaccagccgcaatacaacatgttccaccgttacattgaaaaagaaattatt
cctgtctgcgaacaaaacggcattagccaaatcgttttctcgccgctggcgcaaggggtgctga
ccgggaaatataaacgcgggcaaaaagcgccggaaggaagccgggcgagcgatccgaagtcaaa

tcagtttataaacgacttactaaaagaagaaattcttgctaaagtagaacaactggaaaaagtg
gccgccgagctcggcatcaccctgtcccagctggcgcttgcatgggtattgcgccagccgaacg
tcgcaagcgcgttaattggcgcaagccggccggaacaagtggaagaaaacgtcaaagccgtaga
tgtccaattaacagaagatgtgcttgcgaaaattgaacaaattcttgcatag

MKYRKLGRTGLKVSEISLGSWLTYGNSVEKETAIRVIDKAYELGINSFDTANVYAKGEAEKIVG
EALRKYPRESYVLATKVYWPMGDGPNDRGLSRKHVFEQLHASLKRLQLDYVDIYYCHRYDSETP
IDETLRTIDDLVRQGKVLYVGVSEWTAQQIQEALGTADKYLLDRIVVNQPQYNMFHRYIEKEII
PVCEQNGISQIVFSPLAQGVLTGKYKRGQKAPEGSRASDPKSNQFINDLLKEEILAKVEQLEKV
AAELGITLSQLALAWVLRQPNVASALIGASRPEQVEENVKAVDVQLTEDVLAKIEQILA

7.3.6. ADH E

gtgggaaaagaaatgaatacattctttttgaaaccaaaaatttatttcggaaaccattcattaa
atcatttgtctgattttaatgcaggaaaagtctttattgtaacggatcagacgatgctgaaact
gggcatggcagagaagattatcgaaaaaataaaaggtgctgcgtttaaaatttttccggatgta
gagcctaatccgtccatagaaaccgtcaaaaaggcttttgaatgtttttgcaagaacagccag
agctggtgatagcgcttggcggtggttcagccattgatgctgctaaagcgatgttgctttttta
tcactacatgaaagacatatctgatatagaaatggatttaaaaaaaccattattgattgcaatc
cccacaactagcggaacaggttcagaaatgacatcttattcagtcattacggatacaacgaatc
atttaaaaattcctttgcgtgatgaaaggatgctccctgatgttgccattttagatgagcaatt
aacgataactgtgccaccttctgtcacagcggatacaggcatggatgtgctcactcatgccatt
gaagcatatgtttctttaaactcttcagaatttaccgatatatttgctgagcggtccattaaaa
tggtatttaattatctattaagggcatatcgttttggggaagaccttgatgccagagggaaatt
gcacatagcgtcctgtatggctggtattgcgtttaccaattcgtctttagggattaatcatagc
ctcgcacatgcggttggcgcaaaatttcatttgccgcatggcagaactaatgctattttattgc
cttatgttatccaatataatagcggtctttgcgatgatacgatggatgcttctcctgtggcgaa
gaggtatacagaaatttcgaaaatgttaggcttgccaagctcaaccttaaaagaaggggtcata
agtttggttactgccattcagtttcttaataaaaagctggatataccgtcaagtttcaaagaat
gcgatattaacgaaaccgaatttgcaaaatatatcccttctttggccaaagacgcaatgcaaga
tatttgcacagctggtaatcctagaaaagtaacagaaaaagattttgtctatttattaaaatgg
gcatataacggataa

MGKEMNTFFLKPKIYFGNHSLNHLSDFNAGKVFIVTDQTMLKLGMAEKIIEKIKGAAFKIFPDV
EPNPSIETVKKAFECFLQEQPELVIALGGGSAIDAAKAMLLFYHYMKDISDIEMDLKKPLLIAI
PTTSGTGSEMTSYSVITDTTNHLKIPLRDERMLPDVAILDEQLTITVPPSVTADTGMDVLTHAI
EAYVSLNSSEFTDIFAERSIKMVFNYLLRAYRFGEDLDARGKLHIASCMAGIAFTNSSLGINHS
LAHAVGAKFHLPHGRTNAILLPYVIQYNSGLCDDTMDASPVAKRYTEISKMLGLPSSTLKEGVI
SLVTAIQFLNKKLDIPSSFKECDINETEFAKYIPSLAKDAMQDICTAGNPRKVTEKDFVYLLKW
AYNG

7.3.7. ADH F

atgaaacatctgcaagactgtgtcacactacataacggggtgcaaatgccgtggcttggactcg
gagtgtataaggtgaaagatggagaagaagtcattaatgcggtgcgcactgcgctggaaatcgg
ctatcgccatatcgacacagccgcatattaccaaaatgaagaaggagtcggaaaagcggtccgc
gaatcagggattccgcgcgaagaaatatttatcacgacaaaagtgtggaactctgaccaaggat
atgaaacaacattgaaagcgtttgaaacaagcttaaaaaagttaggccttgattatgtggattt
atatttagtgcactggcccgttaaaggaaaatataaagaaacgtataaggcgctagaaaaattg
tataaagatggacgggtgcgggcgattggcgtcagcaatttccatatccaccatttagaagatt
taatggctgactgcgagattaagccgatggtcaatcaagtggaataccatcctcgtttgacaca
aaaagagcttcatgcgttttgcaagcggcacggcattcagcttgaagcatggtcgccattaatg
agaggggaagttttgcaggaagcggcgcttgttgaaattgggaaaaaatacggaaaaaccccag
cgcaagtcatccttcgttgggatttgcaaaacgaagtggtgacgattccaaagtcggtaacgcc

gcagcgcattaaagaaaacgcagatattttcgattttgagctgacggcggaggaaatggcagca
atcgatgcgcttaatttgaacaaacggatcgggccggatccggacaatttcgatttttaa

MKHLQDCVTLHNGVQMPWLGLGVYKVKDGEEVINAVRTALEIGYRHIDTAAYYQNEEGVGKAVR
ESGIPREEIFITTKVWNSDQGYETTLKAFETSLKKLGLDYVDLYLVHWPVKGKYKETYKALEKL
YKDGRVRAIGVSNFHIHHLEDLMADCEIKPMVNQVEYHPRLTQKELHAFCKRHGIQLEAWSPLM
RGEVLQEAALVEIGKKYGKTPAQVILRWDLQNEVVTIPKSVTPQRIKENADIFDFELTAEEMAA
IDALNLNKRIGPDPDNFDF

7.3.8. ADH G

atgcttactgaaaatatacataagtttgtgaaaaaagagttaattgctttacaggcgaaacaag
tggcagccataatagaagaaaaggaggagttgatttatatggaaaatttttctttttcaaaatcc
gactaaattaatttttgggaaagggcaaattgaacagctgaaggaggaaatcccccgttacggg
aaaaaaatccttctcgtttatggcggcggcagtattaaacggaacggcttgtatgatgaagtaa
tgaacttattgaatgaattaaaggtggaagtaacagaacttcctggcgtggagccaaatccgcg
cttgtccactgtgcggaaaggcgtggaaatttgcaaaacggaaggaattgagttttgctcgcg
gttggcggcggcagtgtcattgattgcacgaaggcgatcgcggcgggtgcgaaatatgatggcg
atccgtgggatttcatcacaaagaaagtgaaagttttcgaagcgcttccgtttgggacagtgtt
aacgcttgcggcaaccggatcagaaatgaacgccggttcggtgattacaaactgggagacgaaa
gaaaaatatggctggaacagtccggcgacgttcccgcaattttcgattctcgatcctacctata
cgctcacagttccgcgggatcatacggtttatggaatcgtcgatatgatgtcccatgtgtttga
acaatatttccatcatacggcgaataccccgctgcaagaccggatgtgtgaggcggtattgcgc
acggtcatggaagcggcgccaaaactagttaatgatttgcaaaactatgagctgcgggaaacga
ttttatattgcggcactatcgctctcaacggcttgttgcaaatggggctccgcggcgactgggc
gacgcataacatcgagcatgcggtgtcagcggtttatgatattccgcacgccggagggctggcg
attttattcccgaactggatgaagcatgtgcttgatgaaaatgtcagccggtttgcccaattag
ctgttcgcgtttttgacgtgaatcctgaagggaaagcggaacgggacgtagcgttagaaggaat
tgaaaaattgcgcgaattttggtcaagcattggggcgccgtcacggcttgccgattacggaatc
ggcgaagaaagtttagaattaatcgctgacaaagcaatggcgaacggcgaatttggcaagttca
aaaagttaaaccgcgacgatgtgctggcgattttgcgcgcatcgttataa

MLTENIHKFVKKELIALQAKQVAAIIEEKEELIYMENFSFQNPTKLIFGKGQIEQLKEEIPRYG
KKILLVYGGGSIKRNGLYDEVMNLLNELKVEVTELPGVEPNPRLSTVRKGVEICKTEGIEFLLA
VGGGSVIDCTKAIAAGAKYDGDPWDFITKKVKVFEALPFGTVLTLAATGSEMNAGSVITNWETK
EKYGWNSPATFPQFSILDPTYTLTVPRDHTVYGIVDMMSHVFEQYFHHTANTPLQDRMCEAVLR
TVMEAAPKLVNDLQNYELRETILYCGTIALNGLLQMGLRGDWATHNIEHAVSAVYDIPHAGGLA
ILFPNWMKHVLDENVSRFAQLAVRVFDVNPEGKAERDVALEGIEKLREFWSSIGAPSRLADYGI
GEESLELIADKAMANGEFGKFKKLNRDDVLAILRASL

7.4. ADDITIONAL SEQUENCES

Following are all gene and protein sequences of additional putative ADHs that were isolated as part of this work, but not further referenced or studied. They will be identified by their internal designations to avoid confusion with previously mentioned enzymes. These sequences correspond with the primers mentioned in Table 35.

7.4.1. ADH 1

atgagtgtttcgcgaattgtatttacatcgctgaactatgtcggctgggggcattggataact
tgcttccggaagtggagcggttttctccgaagaaaatattagtcgtaacagatccagcgttaga
gaaaatcggccttgttcaacgtgtgacagatccgcttgcacaacgtgggtatgacgttcagctt
tatacagatgtagtgccggaaccgccgctggagacgggagaaaaattagtttcctttacgaggg
aaggaaagtttgacctcattatcggcgtcggcggcggaagtgcgatggatttggcgaaactggc
ggcggttttggctgcacatgaagggaaagtagccgattaccttaatttaacgggaacgaaaaaa
gtcgagaaaaaaggcctgccgaaaatattaattccgactacatcagggaccgggtcagaagtga
caaatatttctgtgctctctttagatacgacaaaagatgtggtaacacatgattatttattagc
agacgtggcgattgtcgacccgcaactgaccgtttctgttccgccgcgcgtgacagctgcgaca
ggcattgatgcgctgacccacgcagttgaagcatacgtatccgtcaatgcaagcccaacgtcag
acggtttagcgctgcaggcgattcgtttgatcgcgcggtcgctgcgcaaagcggtggaaaacgg
agcggacaagcaggcgcggatcgatatgagcaacggcagctatttggcgggattggcgtttttt
aacgctggggtcgcaggagttcatgcgctggcctacccgttaggcggacagtttcatatcgccc
acggtgaatcgaatgcggtgttgcttccgtatgtgatgggatatatccgcaaaagctgcacgaa
acggatggccgatattttgaacgcactaggaggaaactcaagctttctctctgaagaagaagct
tcgtataaatgtgtcgaagaactggagagaattgtccgcgatgtcggcattccaagaacgctcg
gtggattcaacatcccagagagcgcattggaaagcttaacgaaagatgccgtccagcaaaaacg
cttgcttgcgcgcagcccgcttccattgcttgaggacgatatacggacgatttaccaatccgct
tttcggggtgtggtaacggagccgaaataa

MSVSRIVFTSLNYVGWGALDNLLPEVERFSPKKILVVTDPALEKIGLVQRVTDPLAQRGYDVQL
YTDVVPEPPLETGEKLVSFTREGKFDLIIGVGGGSAMDLAKLAAVLAAHEGKVADYLNLTGTKK
VEKKGLPKILIPTTSGTGSEVTNISVLSLDTTKDVVTHDYLLADVAIVDPQLTVSVPPRVTAAT
GIDALTHAVEAYVSVNASPTSDGLALQAIRLIARSLRKAVENGADKQARIDMSNGSYLAGLAFF
NAGVAGVHALAYPLGGQFHIAHGESNAVLLPYVMGYIRKSCTKRMADILNALGGNSSFLSEEEA
SYKCVEELERIVRDVGIPRTLGGFNIPESALESLTKDAVQQKRLLARSPLPLLEDDIRTIYQSA
FRGVVTEPK

7.4.2. ADH 3

atgaaaactgcgaaaattgttcagcacaaaagaccgttggaaattttaaatgtcccggatcctc
aaccaggtccggaagatgcggtaattaaaatcgaagcttgcggagtttgccgaagcgactggca
tgcgtggcaaggggattgggcgtggattggtttgtccccggaattgccgataacacctggacat
gaatttgggggtgtcatcgaagaagttggcaaagaggtgaagtcgttccgtccgggagatcgtg
tcaccgtcccgttccattccgcttgcgggcgatgtgaatattgcaaaaaaggtgtgccgaattt
atgcgaaaacctgcaaatttatggcttggtttctggattggaaggcggttatgcggaatatgta
ttagttcgcaatgccgattttaacttgatcagactccctgaaaatgttgacagtttaacagcag
ctgctttaggctgccgttacatgacgggatatcacggtattgtcagaggacatgtcaagcctgg
tgattgggtggctgtgcatggagctggcggtgtaggtctttctgctattcaggtggccaatgca
ttaggagcgcaagtgatagccgttgatattgatgatcaaaagcttgaaatcgcaaaacaggaag
gcgctattgcggccgtcaatgctagaaaagaaaacgttgtcgaagcaataaaggaaatcacaaa
aggggggcgcacatgtcggactggatgctttaggcatcaaagataccgttcttaattcggttcta
tccttaagaaaaggaggaagacatgttcaagtgggtttaaccacatcagaagaaggaggttttg
tctcgctccctgttgacctgattacagcatcggaaattgagtttgttggaagcatcggcaatcc
tcatcctgactatcgcggtttgctgagcttgatttcttccgggcggttgaatccgaagcgctta

gttgaacgtgaaatcaaattggaagaggtaaacgctgtatttgaaaatatgtcacaatataata
cgaaagggtttaacgtaatcactaaattttaa

MKTAKIVQHKRPLEILNVPDPQPGPEDAVIKIEACGVCRSDWHAWQGDWAWIGLSPELPITPGH
EFGGVIEEVGKEVKSFRPGDRVTVPFHSACGRCEYCKKGVPNLCENLQIYGLVSGLEGGYAEYV
LVRNADFNLIRLPENVDSLTAAALGCRYMTGYHGIVRGHVKPGDWVAVHGAGGVGLSAIQVANA
LGAQVIAVDIDDQKLEIAKQEGAIAAVNARKENVVEAIKEITKGGAHVGLDALGIKDTVLNSVL
SLRKGGRHVQVGLTTSEEGGFVSLPVDLITASEIEFVGSIGNPHPDYRGLLSLISSGRLNPKRL
VEREIKLEEVNAVFENMSQYNTKGFNVITKF

7.4.3. ADH 4

atgaaagcggcagttgtcaacgagtttaaacaaaaattagaaattaaagaggtggaaaaaccaa
agctgaattacggagaggtgcttgttaaaattgaggcttgcggcgtttgccacaccgatttgca
tgcggcgcacggagactggccggtcaagccaaagcttcctttgattcctggacacgaaggggta
ggcatcgtcgtcgaggtggcagaaggggtaaaatcggttaaagtaggagaccgtgtcggcattc
catggctgtactccgcttgcggagaatgtgaatattgtttgagcgggcaagaaacgctttgtcc
gcatcaattaaatggcggatattctgccgatggaggatatgcggaatactgcaaagcgcctgcc
aattatgtcgcaaaaattccggaacacttggatccggtggaagtcgcgcctattctttgcgcgg
gggtaacgacatacaaagcgctaaaggtatctaacgccaaaccgggagaatgggtcgccattta
cggaatcggagggttagggcatatcgcccttcaatacgcgaaagcaatgggattaaacgtcatc
gcggtcgatattagcgatgaaaagatagagcttgcgaaacaattaggcgctgatatcgccatca
atggacttaaagaagatcctgtggaagccatccagcaaaacgtgggcggagcgcatgccgccat
tagcgttgccgtcacgaaaaaagcgttcgaacaagcctatcaatccgtgagacgcggcggatgc
cttgttgttgtcggtctgcctaatgaagacttgccaattcctattttcaacacagtattaaatg
gaataacggtgaaaggatccatcgtcggcacgagaaaagatatgcaagaagcgttggatttcgc
cgcgcggggaaaagtgcgcccgatcgtcgaaaccgctccattggaaaaaatcaatgaagtattt
gaaagaatggaaaaaggaaaaattaacggccgagtcgttttaaccatcggtgtcaaccgttaa

MKAAVVNEFKQKLEIKEVEKPKLNYGEVLVKIEACGVCHTDLHAAHGDWPVKPKLPLIPGHEGV
GIVVEVAEGVKSVKVGDRVGIPWLYSACGECEYCLSGQETLCPHQLNGGYSADGGYAEYCKAPA
NYVAKIPEHLDPVEVAPILCAGVTTYKALKVSNAKPGEWVAIYGIGGLGHIALQYAKAMGLNVI
AVDISDEKIELAKQLGADIAINGLKEDPVEAIQQNVGGAHAAISVAVTKKAFEQAYQSVRRGGC
LVVVGLPNEDLPIPIFNTVLNGITVKGSIVGTRKDMQEALDFAARGKVRPIVETAPLEKINEVF
ERMEKGKINGRVVLTIGVNR

7.4.4. ADH 5

atgaaagcgttggtcaaaaaagaattgggttttggcaacctagaaattgtagatgttgtagaac
cagaagtgggaaaagatcaggtcaaaatattagtgaaatatagtgggatatgtggtactgacat
tcacacgtacgaaggtcattataaagtaaaggcacctgttgttttgggacatgaattttcaggg
aaaatcgttgaggtcggtgaaaacgtaacagaatttaaaccgggggacagagtaacgtcagaaa
caactttctatatttgtggggaatgcgtttattgtcaaacaggcgattacaatttgtgtagcag
tagaaaagggctaggaacgcagcaaaatggcagttttgctaagtatgtaattgctagaaaagaa
agtgttcacaagcttcccgaacaagtagattatgtatctgctgctatgactgagccacttgctt
gcgcacatcatgcggtgactaaagcaaccattagcgaaggagatgttgttgtcgttctaggacc
tggcccaattggtttacttgttgctcaagtagtaaaaacttatggagcgactgttattgttgct
ggtttgtctactgatcaacttcggctagaaaaagctaaagaattgggggtggattatgctgtta
atattcaagaagaagacatcaaccatattgtaaaaagtttaacagatggatatggtgctgatgt
tgtctttgaatgcgcaggagctgtttcggcggtgaacatggggctagacttattgaaaaagaaa
gggcagttgatacaggtaggtatttttgcaaaacctgagattcctattaacgcagaaaagatta
ttcaaaaagaaattaaactgattggatctagaagccaaaaacctgctgattgggagccgtcatt
agctttgatgagtgaaaagaaagtaaatgccaaagcgcttgttacgcatcaatttactattgat

caatgggatgaagcgtacagggtcattaaaaatggagaagccatcaaagtattgttaacaccta
tcgattga

MKALVKKELGFGNLEIVDVVEPEVGKDQVKILVKYSGICGTDIHTYEGHYKVKAPVVLGHEFSG
KIVEVGENVTEFKPGDRVTSETTFYICGECVYCQTGDYNLCSSRKGLGTQQNGSFAKYVIARKE
SVHKLPEQVDYVSAAMTEPLACAHHAVTKATISEGDVVVVLGPGPIGLLVAQVVKTYGATVIVA
GLSTDQLRLEKAKELGVDYAVNIQEEDINHIVKSLTDGYGADVVFECAGAVSAVNMGLDLLKKK
GQLIQVGIFAKPEIPINAEKIIQKEIKLIGSRSQKPADWEPSLALMSEKKVNAKALVTHQFTID
QWDEAYRVIKNGEAIKVLLTPID

7.4.5. ADH 7

atggatgttattgctattaccggggccggtacagggttagggaaagaattggcattgcaatatg
ccgggcaaggccatactgtcatcgcccttggcagacggatatcccctcttttagacgtagtgca
ggaaatcaatcagcttggcggaaaagcgtttgcctattcccttgatatccgccgttacgaagat
gttgttaacacagtcgatatgatcgcgaaacatcatcgtgttacgtgtcttgtgaataacgctg
gcatcggtcatttcgggccgctttcttcgctttccgtgcaacaaattgacgatatgattcaaac
gaacgtcaatgggaccatttatatgacgaaagcgtttttgccatattttctgacgatgccaaaa
gcaaaaattatcaatattatttcgactgccggcttacgcggcaaagtgaacgaatccgtctatg
ccgcgacaaaattcgcgattcgcggctttagcgaaagcttgactaaagagctggaaggaacgaa
tatttcggttaccgctgtgtacatgggaggtatggacaccccattctggcatggatcggatcat
ataaaagatcgtttgcggcttcgttctcctaaagaaatagcaaaacaaattatcgcattagaac
atgggcaaaacgaaatgattttatcttaa

MDVIAITGAGTGLGKELALQYAGQGHTVIALGRRISPLLDVVQEINQLGGKAFAYSLDIRRYED
VVNTVDMIAKHHRVTCLVNNAGIGHFGPLSSLSVQQIDDMIQTNVNGTIYMTKAFLPYFLTMPK
AKIINIISTAGLRGKVNESVYAATKFAIRGFSESLTKELEGTNISVTAVYMGGMDTPFWHGSDH
IKDRLRLRSPKEIAKQIIALEHGQNEMILS

7.4.6. ADH 8

atggtcacgaaacagcaggcaacgctgccgccgcagcagcaaaaccgccagcccgggttgcaaa
cggacatgaacccgcagcctgtttcgatcagcgcaacatataaaggaagcgggaaattgaaaaa
taaagctgccatcattagcggcggcgacagcggcatcggccgggcggtcgcgattcattttgca
aaggaaggcgcagatgtggcgatcatttatttaaacgaacatgaagacgcggaagaaacgaagc
gccaagtcgaacaggaaggaagaaaatgcttgctttttgccggcgacgtcggtgatgagcagtt
ttgcaaacatgccgtaaagcaaacgatcgaccaatttggcaaattagacatcgtcgtcaataac
gccgccgagcagcatccgcaaaaaagcttgttgcacattacgtcagaacagctggaaaaaacgt
tccgcaccaacgtcttcggctatttttatttgacaaaagcggcgcttccgtacttgcaaaaagg
cagcgccattattaatacggcgtcaattaccgcttacgaaggcaacgaacaattaattgactat
tcagccaccaaaggcgcgatcgtcgcctttacgcgctcattagcgaaatcgctcgccggtcaag
gcatccgcgtcaacggcgtcgccccgggaccgatttggactccgctgattccgtctacttttac
aagcgaccaagtcgccactttcggctccaacaccccgatgaagcggccggggcagccgagcgaa
gtcgccccaagctacgtctttttggcaagcgatgacgcatcttatattacaggacagatgatac
acgtcaacggtgggaaagtggtgaatgggtaa

MVTKQQATLPPQQQNRQPGLQTDMNPQPVSISATYKGSGKLKNKAAIISGGDSGIGRAVAIHFA
KEGADVAIIYLNEHEDAEETKRQVEQEGRKCLLFAGDVGDEQFCKHAVKQTIDQFGKLDIVVNN
AAQHPQKSLLHITSEQLEKTFRTNVFGYFYLTKAALPYLQKGSAIINTASITAYEGNEQLIDYS
ATKGAIVAFTRSLAKSLAGQGIRVNGVAPGPIWTPLIPSTFTSDQVATFGSNTPMKRPGQPSEV
APSYVFLASDDASYITGQMIHVNGGKVVNG

7.4.7. ADH 9

atgaagttttcacaaaaagtggttgtcgtaaccggcggagcaaatggcattggaaagtcgattg
ctacgatgtttgcggaaaaaggagcgaatgtggtgattgccgatattgctgaagaagcggggaa
gaaacttgttctggcgttaaaagaaaaaggattagaggcgcgctttattccgactgatgttcgg
aaagtcgatgaaatagagcgcctgatgagcgcaacgtttgaaatgtatggcagtttgcattatt
taattaacaatgcgggtgtgtcaagatggaaatctccatatgagttaacggtggaagagtggga
cgatgtgcttcacaccaacttacggagcgtctttttggatcgcgcgaagcagcgaagtatatg
cggaaaaatgaaacaggcggagcgattgtcaatatcgcgtcaacaagggcgctcatgtcggagc
ctcactcggaagcatacgccgcttctaaaggaggaattttgtcattgactcatgcgttagctgt
gtcgtttgccggtgattgtattcgcgtgaacgcaatcagcccgggatggatcgaaacgggggat
tatggccaattgcgggaaattgaccatgcacagcatccagcgcagcgtgtcggcaagccggaag
atatcgcgcgcgcttgtctgtatttgtgcgatgatgaaaatgattttatcactgggacaaacat
tgtcattgatggcggaatgacgagaaagatgatttacgccgagtag

MKFSQKVVVVTGGANGIGKSIATMFAEKGANVVIADIAEEAGKKLVLALKEKGLEARFIPTDVR
KVDEIERLMSATFEMYGSLHYLINNAGVSRWKSPYELTVEEWDDVLHTNLRSVFFGSREAAKYM
RKNETGGAIVNIASTRALMSEPHSEAYAASKGGILSLTHALAVSFAGDCIRVNAISPGWIETGD
YGQLREIDHAQHPAQRVGKPEDIARACLYLCDDENDFITGTNIVIDGGMTRKMIYAE

7.4.8. ADH 12

atgatccgatttgccaccatcggcacaaactggattacggaagcgtttatcgaagcggcaaaaa
agacagaagactttgcgctggccgctgtttattctcggacggacgaaaaagcgaaacaatttgc
ggcgaaaacaggagctgaacggacgtttacaaaccttgaagaactggcgaaaagcaaagacatc
gatgccgtctatatcgcaagcccgaattcgttgcacgccgagcaggcgattttgcttatggacc
acggcaaacatgtcctatgcgaaaagccgatggcgtcgaacacaaaagaagtcaaagcgatgat
cgatgcggcgcgccgcaatggcgtcgttttgatggaagcgatgaaaacgacgctgcttccaaac
tttcaagcggttcgcgaacatttgcataagatcgggaaaatacgccgctattttgcgagctact
gccagtattcgtcgcgctacgacgcgttcaaacaaggaacggtgttaaatgcgttcaatcctgc
cttttccaacggcgcgttaatggatatcggcgtctattgcatttatccgatggtcgtcttgttt
ggaaaaccaaaccgcttgcaagcaagcagcttgaagctcgaatcgggcgtggacggcgaaggga
cgattatctttacttatgaggatatggacgctgttgtcatgtactcaaaaattacgaactccta
tctgcctgtggaaattcaaggagaagacggaagcattctcatcgacgcgatccatacgccaaca
aaagtggaaatccgctaccgcgatgggcgcaccgaagatattaccgtcccgcaagaacaacctc
cgatgtattacgaagtgaaagagtttatcgagctgatcaaaaatggaaaacgcgaatccgaagt
caattcacatgagcattcgttgctgacgatcgctttgatggaagaagcccgcaaacaaacgggc
atcgtatttcctgcagaccaataa

MIRFATIGTNWITEAFIEAAKKTEDFALAAVYSRTDEKAKQFAAKTGAERTFTNLEELAKSKDI
DAVYIASPNSLHAEQAILLMDHGKHVLCEKPMASNTKEVKAMIDAARRNGVVLMEAMKTTLLPN
FQAVREHLHKIGKIRRYFASYCQYSSRYDAFKQGTVLNAFNPAFSNGALMDIGVYCIYPMVVLF
GKPNRLQASSLKLESGVDGEGTIIFTYEDMDAVVMYSKITNSYLPVEIQGEDGSILIDAIHTPT
KVEIRYRDGRTEDITVPQEQPPMYYEVKEFIELIKNGKRESEVNSHEHSLLTIALMEEARKQTG
IVFPADQ

7.4.9. ADH 13

atgcgcaaaaagaaacagattactcgaatgaaaaggaatggggaatgtatcgagtgcgaaaaag
aagaggaggataaagtagtggaacgaattcgtttagcagaagatttaacattttcccgcatcat
tcaaggattatggagactagcggagtgggattattcaaaagaagagatattaagtttcgtggaa
ttttgtctcgaacaagggataacgacgtttgaccatgcggatatttacgggaattatacgtgtg
aagcgttgtttggcgaagcgctggcgttaaagccaagcttgcgcggacaaatcgagattgtgac
aaaatgcgggattaagcttgcctcgaaatacggcttgaagcattatgatacgagtaaggagcat
attatcgcctccgtccaccaatcgttaaaaaatttccgcaccgattacatcgatgtgctgctta
ttcaccgtccggatccgtttatgaatccggaagaagtggccgaggcgttcacccggctgaaaaa
agaaggaaaagtgcgccattttggcgtgtccaattttaaacgttcacagtttaacatgctgcaa
tcgtatttagattttccgcttgtgaccaaccagatagaagtatcagcttactgtttggaaaata
tcgaagacggtacggtcgatttatgtttggaaaagcggattccgccgatggtgtggtccccgct
tgccggtggaaaactatttaccgcgacggacgacagggcggtgcgcctgcgcgcggcgctcgaa
caaattcgcaacaaacttggcgccgaaacgatcgacgaagtattatacgcatggctgcttgtgc
atccggcgcgaatgatgccgattgtcggttccggcaaaaaagaacgcattattcgtgcagtgcg
tgcgctgaagctgccgcttcgccgcgagcagtggtttgaaattttgcaaagctcggtcggccat
gaggtgccataa

MRKKKQITRMKRNGECIECEKEEEDKVVERIRLAEDLTFSRIIQGLWRLAEWDYSKEEILSFVE
FCLEQGITTFDHADIYGNYTCEALFGEALALKPSLRGQIEIVTKCGIKLASKYGLKHYDTSKEH
IIASVHQSLKNFRTDYIDVLLIHRPDPFMNPEEVAEAFTRLKKEGKVRHFGVSNFKRSQFNMLQ
SYLDFPLVTNQIEVSAYCLENIEDGTVDLCLEKRIPPMVWSPLAGGKLFTATDDRAVRLRAALE
QIRNKLGAETIDEVLYAWLLVHPARMMPIVGSGKKERIIRAVRALKLPLRREQWFEILQSSVGH
EVP

7.4.10. ADH 15

atgaaaaaacggcgcatcggtttgtcagatttatatgtcagcgaaatcgggctcggctgcatgt
cgcttggcacggacgagcagaaagcgatccgcattattcatgaagcgctcgaacgcggcattaa
ctatttggacacggccgatctttacgaccgcggattaaatgaagaatttgtcggaaaggcgatc
aaagggaaacgcgatcaagtcatactggcaacaaaagtcggcaaccgctgggaagagggaaaag
acggatggttttgggacccgtccaaaaaatacattaaatcggaagtaaaagaaagcctgcgccg
tctgcaaacagactacatcgacttatatcagctccacggcggcacgattgatgacccgatcgat
gaaacgattgaagcgtttgaagagctaaaacaagaaggggtcatccgctattacggcatttctt
cgattcgtccgaacgtcattcgcgaatatgtaaaacggtcgaacatcgtcagtgtcatgatgca
atacagcttgctcgaccgccgccctgaagaatggtttccgctgctaagcgaacatcacatcagt
gtcatcgcccgcggacctgtcgcgaaaggattgctgacagagcggccgttagaggcggcaagcg
atgctgttaaagaaaacggttatcttgattatacatacgaagaattgcaaacactgattccaaa
gctaaaagcaaaaacagccggacatcgttcattcacagcaacggcgctgcagttttgcttatat
gatccggtcgtcgccgccgccattccaggggcaagcagcatcgaacagatcatcgaaaacataa
acgccgcttccgccccacgtttaaccaaggaagaatatgagtggctccgccagcatacgaaagc
ttcagtgtacaacatgcatcggtga

MKKRRIGLSDLYVSEIGLGCMSLGTDEQKAIRIIHEALERGINYLDTADLYDRGLNEEFVGKAI
KGKRDQVILATKVGNRWEEGKDGWFWDPSKKYIKSEVKESLRRLQTDYIDLYQLHGGTIDDPID
ETIEAFEELKQEGVIRYYGISSIRPNVIREYVKRSNIVSVMMQYSLLDRRPEEWFPLLSEHHIS
VIARGPVAKGLLTERPLEAASDAVKENGYLDYTYEELQTLIPKLKAKTAGHRSFTATALQFCLY
DPVVAAAIPGASSIEQIIENINAASAPRLTKEEYEWLRQHTKASVYNMHR